THE UNIVERSITY *of* EDINBURGH

Foundational Technologies in Synthetic Biology: Promoter Measurement and Peroxisome Engineering

Kim Stephen de Mora

A Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2010

Abstract

The confluence of next generation DNA sequencing and synthesis when combined with the application of concepts such as standardization and modular design has led to the genesis of a new discipline. The nascent field of Synthetic Biology concerns the rational design and construction of genetic circuits, pathways, machines and eventually whole organisms. The immaturity of this field dictates that early research efforts, including this Thesis, describe foundational work towards the creation of tools which make biology more amenable to being engineered.

The first part of this Thesis describes an attempt to standardize the measurement of transcriptional promoter activity in *E. coli*. A method to measure *in vivo* promoter activity was developed for *E. coli* and tested in a multi-institution trial. Comparable results were achieved with less than a two-fold range for the measured promoters across eight laboratories. A standardized measurement kit was created and distributed for use by the teams participating in the 2008 international Genetically Engineered Machines competition. Techniques learned measuring the activity of *E. coli* promoters were applied to a collection of *S. cerevisiae* strains. Several promoters were measured in synthetic dextrose media and *ADH1* was measured in multiple media conditions. The outcome of these experiments is to consider proposing *ADH1*as the reference promoter in *S. cerevisiae*.

The second aspect of this Thesis describes the construction of artificial organelles in *S. cerevisiae*. Artificial organelles hold the prospect of being able to insulate synthetic genetic pathways from the cell. Two proteins are essential for the biogenesis of the peroxisome organelle in humans and yeast, Pex3p and Pex19p. Pex3p functions as a peroxisomal membrane receptor for Pex19p, while Pex19p shuttles other peroxisomal proteins to the membrane, including Pex3p, creating a feedback loop. Human Pex19p has previously been shown to dock to yeast Pex3p and a version of yeast Pex19p has been shown to work with human Pex3p as a high degree of evolutionary conservation exists between these proteins. Because of these inter-species protein docking characteristics, there exists the possibility of creating bimodality: the ambition of the work was therefore to create a cell strain which possessed both synthetic “humanized” and natural yeast peroxisomes. An *S. cerevisiae* BY4741a derivative strain was engineered with fluorophore tagged versions of human (CFP) and yeast (YFP) Pex3p and untagged yeast and human Pex19p proteins. The results indicated the creation of a single population of peroxisomes when a measure of fluorescently imaged CFP and YFP peroxisomes were plotted on a scatter plot. A log of the ratio of CFP to YFP peroxisomes was plotted on a histogram and a normal distribution was found to best fit the curve, indicating a lack of bimodality. Finally, microscopy images of this strain were reviewed and by visual inspection, showed no evidence of distinct human or yeast peroxisomes. This experiment therefore produced no evidence of bimodality when examining the interactions of human and yeast Pex3p and Pex19p proteins. However, the four proteins were shown to interact closely to produce a single population of chimeric human-yeast peroxisomes. The peroxisome-deficient mutant phonotype strain was rescued using human Pex3p and Pex19p.

Declaration

I declare that this Thesis has been composed by myself and is all my own work except where otherwise stated.

Signed

Kim Stephen de Mora
Edinburgh 2010

To My Friends and Family,

I would like to thank my family for help, advice, support, money and most importantly, tolerance. PhD research is easy until you have to write about it. For help during this difficult time, I am eternally grateful to my Mother, Father and my little brother Lee. Who still completed his PhD before me.

I also want to thank long suffering friends Bill Edwards, Martin Lee, Jovana Radulovic, John Watson, Agata Daleszynska, Hugh Sturgis, Mackenzie Cowell, Natalia Bednarska, Kristen Albertsen and Em Withnell. The end of this process was difficult for me. I've tested their patience many a day and yet they still speak to me. For this I am also grateful.

Kim de Mora

Acknowledgements

I would like to acknowledge the help of my supervisor, Dr Alistair Elfick, for all the conversations on Synthetic Biology and the numerous opportunities for ‘networking’ as well as my second supervisor, Dr Kevin Hardwick. I would like to thank Professor Pam Silver for hosting me in her lab at Harvard Medical School. Without her help and the help of her students, this work would not have been possible.

Dr Chris French has been of enormous help in showing me the basics in molecular biology and teaching me good sterile technique. Bruno Afonso was of great help in learning the basics of fluorescence microscopy using the Nikon microscope. Dr David Kelly was invaluable in learning how to use the DeltaVision microscope, in programming the Image Pro script and with numerous discussions about the results from both the Nikon and the DeltaVision microscopes. Finally, thanks to Jake Wintermute for teaching almost everything I know about molecular biology and for giving me this project.

Table of Contents

Table of Figures

Table of Tables

Glossary of Biological Terms and Abbreviations

BBF	BioBricks Foundation
BBSRC	Biotechnology and Biological Sciences Research Council
BiFC	Bimolecular Fluorescence Complementation
BPP	Bits Per Pixel
CIP	Calf Intestinal Alkaline Phosphatase
COIL	Centre Optical Imaging Facility (University of Edinburgh)
CPRG	Chlorophenol red-β-D-galactopyranoside
CSAIL	Computer Science and Artificial Intelligence Laboratory
DIC	Differential Interference Contrast (microscopy)
ELSI	Ethical, Legal, and Social Issues
EMCCD	Electron-Multiplying Charge Coupled Device
EPSRC	Engineering and Physical Sciences Research Council
ER	Endoplasmic Reticulum
ESRC	Economic and Social Research Council
FACS	Fluorescence Activated Cell Sorting
FBI	Federal Bureau of Investigation
HMS	Harvard Medical School
HTS	High Throughput Sampler (module)
ImPro	Image Pro software
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
LB	Luria Broth
MIT	Massachusetts Institute of Technology
MM	MetaMorph software
MTMs	Myotubularins
NEB	New England Biolabs
nt	Nucleotide

OD	Optical Density
ONPG	O-nitrophenyl-β-D-galactopyranoside
Peroxin	Peroxisomal Protein
PEX	Peroxin (peroxisomal protein)
Pex3p	Peroxisomal membrane protein in yeast and humans
Pex19p	Peroxisome shuttle protein in yeast and humans
PMP	Peroxisome Membrane Protein
PoPS	Polymerase Per Second
PSF	Point Spread Function
PTS1	Peroxisome membrane signaling protein 1
RFC	Request For Comment
SB	Synthetic Biology
SD	Synthetic Dextrose (media)
S-Gal	Synthetic Galactose (media)
S-KAc	Potassium Acetate (media)
S-YNB	Synthetic Yeast Nitrogen Base (media)
UCSF	University of California, San Francisco
USD	United States Dollar
UN	United Nations

1 Introduction

In 1828, Friedrich Wöhler synthesized urea, an organic chemical compound, from inorganic substances (Wöhler 1828). This act ushered in the dawn of synthetic chemistry and by the 1850's, dyes and pigments were synthesized which led to a revolution in the textile industry as toxic natural compounds such as arsenic were no longer used. By the end of the 19th century, many more organic compounds had been created in the lab. This led to the creation of the petrochemical industry when synthetic rubbers and adhesives were produced from crude oil and its derivatives. After the petrochemical industry, came the pharmaceutical industry where more and more complex organic molecules and proteins were synthesized to create drugs. Drugs like aspirin, insulin and penicillin were manufactured in large quantities leading to a revolution in medicine and healthcare (Gerhardt 1853; Bliss 1982; Pontikis 2010).

The structure of Deoxyribonucleic acid (DNA) was elucidated in 1953 by Watson and Crick (Watson and Crick 1953); a discovery that paved the way for genetics and genetic engineering. The 1968 discovery of restriction enzymes in *Haemophilus influenzae* (Smith and Wilcox 1970) allowed scientists to cut DNA at specific locations, which introduced the possibility of recombinant DNA and the construction of plasmids *in vitro* (Cohen, Chang et al. 1973). The work of Frederick Sanger in 1975 produced the dideoxy sequencing or chain termination method of DNA sequencing (Sanger, Nicklen et al. 1977). The discovery of Taq polymerase in 1976 led Kary Mullis to invent polymerase chain reaction (PCR) technology in the early 1980's (Mullis, Faloona et al. 1986). The advent of shotgun sequencing led to *H. influenza* being the first free living organism to have its entire genome sequenced (Fleischmann, Adams et al. 1995). The convergence of these technologies and advances was leading to one of the greatest scientific endeavors of the 20th century.

The human genome project is the largest biology project undertaken to date. It began in 1990 when James Watson was the head of the National Center for Human Genome Research at the National Institutes of Health in the United States. The USD $3 billion dollar project was

roughly expected to take an international consortium of researchers from the US, the UK, Japan France, Germany and China 15 years to complete a consensus sequence of the haploid genome. The human genome was first published in 2001 when a physical map of the largest genome ever sequenced was published in nature (McPherson, Marra et al. 2001). The same month in Science, Craig Venter and his team published a 2.91 billion base pair consensus sequence of the human haploid genome (Venter, Adams et al. 2001). In 2006, the public consortium project was finally finished after 16 years as the annotated sequence of chromosome 1 was published (Gregory, Barlow et al. 2006). In parallel to the International Human Genome Sequencing Consortium (IHGSC), Craig Venter was the first person to sequence a diploid human genome of an individual (his own) in 2007 (Levy, Sutton et al. 2007). At the time of publishing, the cost of sequencing a human genome ranges from about $5000 USD (Lauerman 2009) to $20,000 USD (Karow 2010) and is expected to fall to less than USD $1000 or cheaper (Johnson 2010) within a few years.

In 2008, Craig Venter published a protocol on how to synthesize and assemble a 582,970 base pair *Mycoplasma genitalium* genome (Gibson, Benders et al. 2008). Due to problems transforming this DNA into *M. genitalium*, a 1.08 mega–base pair *Mycoplasma mycoides* genome was constructed using the previously described method. In 2010, Venter succeeded in inserting the artificial *M. laboratorium* DNA into a DNA-free cell and 'booting up" life (Gibson, Glass et al. 2010). However, his project copied the DNA of an existing species with the addition of a few watermarks in order to test the method. He did not design his own organism from scratch. Costs have not fallen in gene sequencing as much as they have in synthesis and the biggest problem facing synthesis is still the error rate.

The industrial revolution was based on several well known innovations, such as the steam engine, the mechanization of industrial processes, the discovery and mining of coal in the British midlands and one less well known factor: standardization. Factories produced all the parts they needed for their machines, and every factory had a different standard and size for nuts and bolts. The standardization of fasteners allowed factories to concentrate on producing their products, not each individual component. Many examples of standardization exist from the beginning of the industrial revolution to the emergence of the Internet, from the Ohm allowing trans-Atlantic

telegraph cables to be laid and repaired to TCP/IP protocols allowing computers to communicate.

Biology in its current guise is highly standardized in that four nucleotides form the basis of all known living organisms. With 20 amino acids, the structure of all proteins in the biological world can be formed (but protein function is more complex as secondary structures are usually needed for the protein to fold correctly and often metal ions are necessary, specifically in the case of metalloenzymes). The field of biology is poised to go through great changes that will seek to exploit and capitalize on this underlying standardization as it moves further into the 21st century. The nascent field of Synthetic Biology is about the construction of genetic circuits, pathways, machines and eventually whole organisms. The confluence of next generation DNA sequencing, synthesis and the application of engineering concepts such as standardization, decoupling and abstraction (Endy 2005) to biology will pave the way for Synthetic Biology, or SynBio to likely be the next industrial revolution.

1.1 What is 'Synthetic Biology'?

This question can be answered differently depending on your background and interests. There are several groups and many definitions for this new field. First the groups: biologists, bioinformaticians/systems biologists, chemists, re-writers and engineers (Endy 2005). Biologists are interested because they will be able to use new tools and techniques to probe current understanding. They can build systems to test their understanding of the underlying phenomena. Systems biologists and bioinformaticians want to test computer science techniques and *in-silico* models against synthetically constructed biological systems. With biochemistry being an extension of chemistry, Synthetic Biology gives chemists the ability to create new molecules and systems to further probe questions from simple biomolecular interactions to the origins of life. 'Re-writers' seek to optimize biology for human interaction by making biology easier for humans to use through projects such as genome refactoring (Kuldell and Lerner 2009). Finally, for engineers, biology is technology (Carlson 2010). Engineers want to build on the last four decades of progress in genetic engineering by turning biology into a discipline where new functions can be easily created through application of engineering methodologies. In this Thesis, Synthetic Biology will be discussed from the engineering perspective. It is about the creation of tools to make engineering biology easier. However, engineering and biology are

vastly different in the ways knowledge can be used to design systems. Engineering applies theories and material properties to design systems to a specification that can be achieved in a range of operating conditions. The same principles can't be applied to the design of biological systems as the behavior of each individual genetic part can't be predicted in the same deterministic manner as in engineering. This doesn't mean biology can't be engineered, but has implications in the design of the project. A parts based assembly approach is taken in this Thesis, as will be discussed later, but there is scope to develop a better method. In a talk on Synthetic Biology at the Singularity University, Andrew Hessel gives the following definition of Synthetic Biology:

> "An ultimate dream is to design… , press a button and have the design translated into DNA sequences that can synthesized and put to work in living cells (Hessel 2009)"

This quote reflects the ideal situation in a future where cloning is no longer necessary to assemble DNA fragments.

In 1978, the Nobel Prize in physiology or medicine was awarded to Arber, Nathans and Smith for the discovery of restriction endonucleases. In an editorial to the journal Gene, Waclaw Szybalski was the first to term this new field:

> "The work on restriction nucleases not only permits us easily to construct recombinant DNA molecules and to analyze individual genes, but also has led us into the new era of Synthetic Biology where not only existing genes are described and analyzed but also new gene arrangements can be constructed and evaluated (Szybalski and Skalka 1978)."

Szybalski was a little ahead of his time, but restriction enzymes still had a massive impact on the field of biology. The field he called Synthetic Biology became Molecular Biology and the first modern reference to "Synthetic (Cell) Biology" in the literature came in 2001:

> "Molecular Biology has entered a stage of maturity that requires its transformation into an engineering discipline. The wealth of data on cellular components and their interactions will promote an understanding of cellular behavior that is sufficient for prediction, control and redesign (Arkin 2001)."

As the above quote illustrates, the nascent field of Synthetic Biology is about applying a quantitative approach to an often qualitative field. Most work in Synthetic Biology takes place at the transcriptional level. Transcriptional logic has been well understood and used in molecular biology for some years (Casadaban 1975). The process by which DNA has been extracted, copied and written is generally referred to as cloning. Until very recently and still in many labs, building and sharing genetic constructs by cloning was analogous to blacksmithery and cannot be performed reliably and repeatedly with a 100% success rate. The exact restriction enzymes that were necessary for the DNA fragment were built into the PCR primers and each set of restriction enzymes were specific to each construct. This was advantageous in that site directed mutagenesis was seldom necessary, but problematic in that constructs could not be easily shared or reused. Another aspect of Synthetic Biology uses a new and novel construction technique to quickly and easily assemble multiple genetic parts together to produce a device:

> "The field is in a similar situation to mechanical engineering in the 1800s and micro-electronics in the 1950s, when rapid progress required the adoption of standardized interchangeable parts and modular construction methods." (Haseloff and Ajioka 2009)

These devices can be assembled together in a modular manner to build systems to "reprogram" the host organism for a variety of different tasks. Although genetic engineering has existed for decades, conventional techniques were inflexible to producing larger constructions. This simple but powerful new technique is called the "BioBrick" system. This format of part exchange is supported by the BioBricks foundation in an effort to help mediate the technology and inspire students to become involved in the field (Smolke 2009).

The BioBrick system has had a huge impact on Synthetic Biology by allowing laboratories across the world to use standard parts to construct devices and systems. Biological parts can be selected from a catalogue and ordered, in the same way an engineer would select fittings or components. Labs across the world are developing new BioBrick parts and beginning to characterize existing ones. Once submitted to the open source Registry of Standard Biological Parts (Rettberg 2010), a part could be obtained from the registry by simply ordering the whole catalogue and re-hydrating the construct on arrival. This system draws many parallels to open source software engineering. It is a concept that has the possibility to revolutionize biology.

1.2 Exemplars in Synthetic Biology

The field of Synthetic Biology has grown explosively over the last several years and covering all of the representative projects would be impossible. Several illustrative projects or exemplars will be briefly presented to show the milestones in the field.

Arguably, the first demonstrated exemplar in Synthetic Biology was the repressilator. This system was composed of three transcriptional repression systems and was designed to oscillate a GFP signal in *Escherichia coli* over time. The first repressor protein, LacI, inhibits the transcription of the second repressor gene, *tetR*, whose protein product in turn inhibits the expression of a third gene, *cI* from λ phage (Elowitz and Leibler 2000). They succeeded in creating a strain of *E. coli* that would periodically oscillate, inducing the synthesis of GFP. As the period of oscillation was slower than the generation time of the cells, the state of the oscillator was passed from generation to generation. The period of oscillation of the system depends on several parameters: the dependence of the transcription rate on repressor concentration, the translation rate and the decay rate of the mRNA. Strong promoters coupled to efficient ribosome binding sites, comparable mRNA and protein degradation rates and cooperative repression characteristics were necessary for the system to function. The period of the oscillations was in the order of 160 ± 40 minutes.

Another interesting device is the genetic toggle switch (Gardner, Cantor et al. 2000). Gardner and colleagues built a bi-stable toggle switch using two repressors and two constitutive promoters. They designed and modeled the switch based on the biochemical rate equation for gene expression. A simple Ordinary Differential Equation (ODE) model was constructed in Matlab and they then built and demonstrated that their system exhibited bimodality *in vivo*.

There are exemplars that demonstrate other important functions in biology, such as cellular memory (Ajo-Franklin, Drubin et al. 2007). Ajo-Franklin and co-workers built a set of transcriptional activators and quantitatively characterized their effects on gene expression in *S. cerevisiae*. They correctly predicted the behavior of the memory network based on quantitative characterization of the promoter pairs. Simple memory in single-cell eukaryotes is a powerful tool if it can be easily exploited.

There have been several projects from the international Genetically Engineered Machines (iGEM) competition that can also be classified as exemplars. In 2005, a team from Texas produced a strain of bacteria that was light-sensitive, making "bacterial photography" (Levskaya, Chevalier et al. 2005). This was the first truly successful demonstration of an iGEM project. In 2006, the University of Edinburgh team produced a strain of *E. coli* that could detect arsenic to a concentration of 5 ppb or greater than the WHO detection standard of 10 ppb (Aleksic, Bizzari et al. 2007). This project has been championed by the creators of the competition and is still mentioned as a model Synthetic Biology project. From 2006 onwards, iGEM projects started to tackle more complex projects involving mammalian cells. The 2006 team from Slovenia engineered a method of eliminating sepsis in mammalian cells. The binding of bacterial components to a family of Toll-like receptors activates the immune system, where an exaggerated response can lead to systemic inflammation. Mathematical modeling of the system indicated that a feedback loop at the weak spot in the signaling cascade at MyD88 could inhibit this response (Ciglič, Fekonja et al. 2007). In 2008, a team from the same institution tested functional 'immunobricks' by assembling a vaccine to combat acquired immune response to *H. pylori* and tested their system *in vivo* in mice (Čeh, Kočar et al. 2008).

The first commercialization examples of Synthetic Biology are starting to appear. Jay Keasling and his team engineered a strain of yeast to produce the anti-malarial drug precursor artemisinic acid to approximately 50 mg per liter of saturated culture (Ro, Paradise et al. 2006). Today, they have increased that figure to about 120 mg / L of saturated culture. The project was funded by the Bill and Melinda Gates foundation with the aim of producing the drug at much lower cost than current processing techniques, effectively allowing developing countries that needed a cheap supply to purchase enough treatments to inoculate their population. An added bonus of this project was that the artemisinic acid biosynthesis pathway is very close to a class of organic chemicals called terpenoids (isoprenoids), which make Excellent biofuels (Keasling and Chou 2008). This discovery led to the formation of Amyris Biofuels with the aim of producing bio-diesel, bio-jet fuel and bio-gasoline from engineered yeast. Professor Keasling's projects are impressive, but the work took 150 man-years and $25 million USD for the artemisinic acid biosynthesis pathway alone. This scale of work is possible with large grants or follow-up

industrial applications, but not without these huge sources of finance. Keasling's initial work involved many years of developing and creating the tools necessary to make his achievements possible and he has stated with the experience they now possess, the same feat could be reproduced in a much shorter time span with less financial resource.

1.3 Ethical, Legal and Social Issues surrounding Synthetic Biology (ELSI)

The field of Synthetic Biology has the potential to make great changes with wide ranging impacts on people, communities, economies and even nature itself; issues that reach far beyond the confines of the lab. As the implications of this work can reach so far with major discoveries in the field making headlines, discoveries within context become important:

> "Indeed, the 'I' of ELSI itself implies that once the natural scientists have done their work, the social scientists arrive to explore the 'implications' of the work for society, perhaps by drawing analogies with similar technological developments in the past" (Calvert and Martin 2009).

In this quote, Jane Calvert states that the traditional methodology of the social sciences is to investigate the implications of natural scientists work after the 'fait accompli'. In this context, the umbrella of social science refers to the Ethical, Legal, and Social Issues (ELSI) surrounding the field of Synthetic Biology. Social scientists usually take the role of external observers and report on the implications of discoveries after scientists have finished and published their work. The first wave of Synthetic Biology research institutes have incorporated ELSI into the core of their work. The Synthetic Biology Engineering Research Center (SynBERC) has a human practices thrust that aims to foster mixed teams of human scientists with other researchers to design models of collaboration and inquiry (Endy 2010).

In the UK, the Economic and Social Research Council (ESRC) has established a national 'Genomics Network' of research centers to carry out social science research on the science and technology of genomics, with Synthetic Biology being one of the specific areas of investigation (Frow 2009). The Royal Academy of Engineering and the UK Research Councils have commissioned reports investigating the scope, applications and implications of Synthetic Biology as well as several public dialogues to explore peoples informed and uninformed

perceptions (Kitney 2009; Paterson, King et al. 2009). The BBSRC have also published their own reports on the findings of these studies with a preliminary indication that the public finds Synthetic Biology technologies both exciting and scary (Balmer and Martin 2008; Bhattachary, Calitz. et al. 2008). In this field, the role of social scientist is changing and moving away from the traditional impartial observer to a more active role in the development of the field itself. Social commentary has become involvement, a new paradigm for the social sciences. This change is likely to lead to closer relations between the social and physical sciences and where there is collaboration and understanding, productive and positive interactions are much more likely to result.

A prominent ethical issue raised about Synthetic Biology relates to the wider implications of trying to engineer DNA as a programming code. Craig Venter has succeeded in booting up *M. laboratorium* from entirely synthetic man-made DNA (Gibson, Glass *et al.* 2010). The question that then needs to be asked is: "Is this creating life?" In radio interview on the program Frontiers – Synthetic Biology, a social scientist and Anglican priest, Robin Gill, said:

> "I think it's a metaphor [on the idea of creating life], it's not my understanding of creating. For me, creating is creating from nothing. I think what we're doing here is modifying life, combining synthetic elements with bio-elements.[1]"

The issue can be transformed into a question of using a bottom-up or a top-down approach to constructing biology. A top-down approach focuses on modifying existing organisms by changing part of the whole genome. One ongoing effort in this area is Jef Boeke's *S. Cerevisiae* 2.0, where a large international group of collaborators are trying to re-build a minimal genome version of the most widely used experimental yeast (unpublished work). The bottom up approach to Synthetic Biology involves building an organism from component parts. This approach differs as aside from Robin Gill's idea on the creation of life stated above, there are still ethical issues involved. Work in this category would include Jack Szostack' self replicating artificial vesicles (Hanczyc and Szostak 2004).

[1] August 2009, BBC Radio 2

An interesting aspect of the field from a legal standpoint is the issue surrounding intellectual property. Parts-based Synthetic Biology functions with the aid of the Registry of Standard Biological Parts at MIT (Rettberg 2010). This DNA repository exists solely due to the academic exemption to IP law (known as the Hatch-Waxman exemption in the US), where academic institutions can use protected patents without fear of legal pursuit from the patent owners. Many of the parts in the registry are known to be or suspected to be patented including many of the green fluorescent proteins (D Grewal, personal communication). As a consequence, parts-based Synthetic Biology could not exist without this exemption. Questions surrounding the transition into commercialization then become more interesting, and more of a debate. As Synthetic Biology grapples with the issues surrounding IP, a landmark ruling was made in the Association For Molecular Pathology *et al.* v. United States Patent and Trademark Office *et al.* case involving a US company. The plaintiffs' in this case posed a unique question:

"Are isolated human genes and the comparison of their sequences patentable?"

This question brings attention to the intersection of Molecular Biology and patent law. Myriad Genetics own the patents related to specific sequences implied in breast and ovarian cancer in the *BRCA1* and *BRCA2* genes and hold a monopoly on testing these genes for selected mutations. The defendants won the case when parts of the company's patents were ruled invalid in the U.S. District Court for the Southern District of New York (Sweet 2010). In the context of Synthetic Biology, the legality of distributing, using and creating BioBrick parts has remained a grey area. In an attempt to tackle this issue, the BioBrick foundation drafted a public agreement pertaining to the use and distribution of BioBrick parts (Endy, Grewal et al. 2010). This framework simply states that you may use the parts, but in the event that a company seeks compensation, the BioBricks Foundation is not the legally responsible entity.

From the social standpoint, the field is rapidly evolving new facets with the use of community organizing tools. OpenWetWare is a community based web site that allows synthetic biologists to share protocols and useful techniques (Kelly 2010). It evolved from the use of wikis as the main public form of information dissemination for the iGEM competition where producing a team wiki is a key element of the competition (Rettberg 2010). The field has introduced to

biology a new format of publication from Computer Science – Request For Comments (RFC) which will be discussed later in the Thesis.

One of the other aspects of the social science in Synthetic Biology involves the issue of safety, especially in the context of public perception. While the founders of the field are pushing ahead with the technical development to allow discoveries to become possible, there are other opinions being expressed, including from the security community. Synthetic Biology aims to makes biology easier to engineer which in turn makes biological weapons easier to engineer. In 2004 Steve Kurtz was made painfully aware of how the security community can react after being arrested on suspicion of bioterrorism. Paramedics were called to his home when his wife was in cardiac arrest. The art professor was using genetically modified bacteria as part of an art installation for the Massachusetts Museum of Contemporary Art but the paramedics became suspicions of the biological materials in his home and alerted the FBI (Hirsch 2005). He was eventually cleared of all charges after a four year legal battle (Galarneau 2008).

In 2006, a reporter from the Guardian newspaper in the UK succeeded in ordering an partial sequence of the smallpox virus from a company online and had it delivered to his home address (Randerson 2006). The reporter was aiming to write an expose on how lax laws allow new DNA synthesis companies to send sequences without verification of the customers or the sequences. With the widespread availability of genetic information about most of the worst mass pandemic diseases easily found online the idea was that more controls should be in place. While potentially worrying, the 78 base pair sequence ordered represents a tiny fraction of the 185 000 nt smallpox genome. Assembly of the approximately 2400 78 nt oligomers in to a functional version of the virus is not trivial. The Synthetic Biology community is aware of potential biosecurity threats of large-scale DNA synthesis and is trying to work with governments and industry to implement safety procedures (Bugl, Danner et al. 2007). The do it yourself biotechnology (DiyBio) community are also aware of potential threats and have been in discussions with the FBI (Weapons of Mass Destruction Directorate) and the UN (Biological Weapons Convention Implementation Support Unit) to help address these issues (Personal communication M. Cowell, J. Bobe, E. You, P. Millet).

Scientists involved in the field of Synthetic Biology need to be very aware of the implications of what they are doing. Questions pertaining to the release of genetically engineered organisms into the environment are still as relevant now as when GM crops first appeared in the media. Academic research can often proceed in the confines of academia without attracting the attention of the public, but in this case researchers would do well to be aware of broader social context in which Synthetic Biology is developing. The Asilomar conference in 1975 described a set of principles to follow for any researcher working on recombinant DNA (Berg, Baltimore et al. 1975). These principles state that containment be made an essential consideration in the experimental design and that the effectiveness of the containment matches the estimated risk. These principles help guide genetic engineering and are still as valid today for Synthetic Biology as they were when originally conceived.

1.4 The International Genetically Engineered Machine Competition

The main training ground for aspiring Synthetic Biology is an undergraduate competition that takes place every year at Massachusetts Institute of Technology (MIT) (Brown 2007). In 2002, three engineering professors chose to break with the MIT traditions of computers, AI, electronics and robotics at the Computer Science and Artificial Intelligence Laboratory (CSAIL) department to think about how to engineer biology. In the MIT Independent Activity Period in January 2003, they decided to give the challenge of building biological machines to improve on the repressilator (Elowitz and Leibler 2000). It became iGEM in 2005 coinciding with the participation of the University of Cambridge along with 11 US teams. In 2006, 13 teams became 37 from the UK to India to South America. In 2009, the competition had 112 participating teams and it continues to grow year on year. The competition is structured around a team of undergraduates working for a period of 10 – 12 weeks over the summer to design, build and test a genetic machine. The iGEM competition started with the aim of answering a question posed by Randy Rettberg:

> "Can simple biological systems be built from standard, interchangeable parts and operated in living cells? Or, is biology simply too complicated to be engineered in this way?"

After a national level competition and two years of international work, the question was answered in the affirmative as teams demonstrated working biological devices at the 2006 final (Aleksic, Bizzari et al. 2007; Ciglič, Fekonja et al. 2007; Kitney, Freemont et al. 2007). The competition is currently the most well-known forum for the generation of new Synthetic Biology project ideas and the focal point of the international research community. It is also how the author became interested in the field.

1.5 The Registry of Standard Biological Parts and BioBricks

The Registry of Standard Biological Parts is both a database of parts and a catalog of DNA that can be ordered. It was established in 2003 to provide support for the iGEM competition. There are many categories of parts available, the most common being promoters, ribosome binding sites, genes (especially fluorophores) and terminators. Most of the documentation to build and use BioBricks can also be found on the site (Rettberg 2010). The principal model organism for Synthetic Biology is *E. coli* as nearly all genetic assembly is performed in the bacteria. The assembly of the parts is performed in *E. coli* and as a consequence, most of the parts in the registry are for this organism. DNA assembly and cloning for most other organisms also takes place in *E. coli*.

The distribution of parts is traditionally available through participation in the iGEM competition although it can be ordered by any interested lab. At the time of writing, about 1200 parts are distributed by the registry each year. The parts are submitted by iGEM teams as a requirement of the competition. They are sequenced by the teams but little characterization work is usually performed to determine the characteristics of the parts beyond functionality. Many of the parts described in this Thesis are available in the Registry of Standard Biological parts. Traditionally, the parts are distributed in 384 well plates but in 2008, the registry was spotted onto chromatography paper and sent in a three ring binder. This format was copied for the *E. coli* promoter measurement kit that described in Chapter 4.

The BioBricks Foundation has also implemented a publication system from computer science that allows for the open discussion on the setting of standards. This system draws heavily on the MIT origins of software engineering and is called the Request For Comments (RFCs) (Knight

2007). This forum is where discussions take place on the setting of standards for the community, including the assembly standards and the electronic BioBrick description standards.

The BioBricks™ technology was first described by Tom Knight at MIT (Knight 2003). The technology describes a system of assembling fragments of DNA together from a set of standard interchangeable parts. BioBricks are defined as genetic parts that contain standard restriction enzymes flanking a piece of DNA of interest. The first standard, initially known as assembly standard 1.0 (now known as RFC10) contains four standard restriction enzymes: *EcoRI*, *XbaI*, *SpeI* and *PstI*. *EcoRI* and *XbaI* are placed upstream of the part and *SpeI* and *PstI*, downstream. A Not1 site was placed between the *EcoRI* and *XbaI* sites as well as the *SpeI* and *PstI* sites to allow enough space for the enzymes to bind. The following diagram explains BioBrick single antibiotic standard assembly:

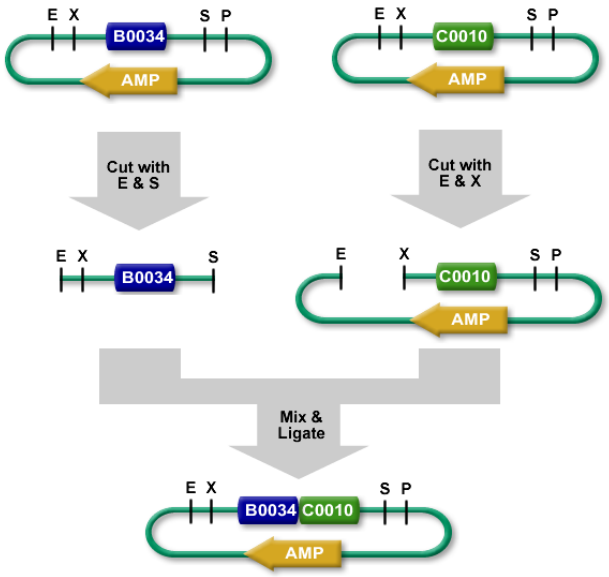


Figure 1.1: BioBrick assembly diagram (Rettberg 2006)

Figure 1.1 shows BioBrick part assembly in diagrammatic form. The blue part is inserted downstream of the green part when cutting the fragments with the given enzymes. It is also possible to insert the blue part upstream of the green part by cutting it with *XbaI* and *PstI*, while the green part (vector) would be cut with *SpeI* and *PstI*. A variant of this assembly technique can be used to assemble three parts together in one step called three antibiotic assembly. This

two part assembly technique was used to assemble all constructs involved in the projects discussed in this Thesis.

When the original BioBrick assembly standard was first described, a limitation in the design was overlooked. This original standard is now referred to as BBF RFC assembly standard 10. Once the parts have been assembled together, translation over the mixed assembly site causes a frame-shift of one nucleotide, preventing the assembly of fusion proteins. This issue was addressed with the creation of new BioBrick assembly standards. There are currently six BioBrick assembly standards in the Registry: 10 (Knight), 12 (Knight 2), 21 (Anderson), 23 (Silver), 25 (Mueller) and 28 (Peisajovich). These standards have been created to address a range of issues from fusion protein creation to problems with charged residues in the mixed assembly site. As eukaryotic organisms are incompatible with polycistronic expression, a new assembly standard was needed to allow eukaryotic Synthetic Biology to proceed. The Silver-Philips standard was developed in Pamela Silver's lab to allow the construction of fusion proteins (Phillips and Silver 2006). This new standard simply removed a single nucleotide from the RFC10 to allow the correct frame to be preserved when translating over the *XbaI-SpeI* mixed site. All BioBricks used in the yeast work in this Thesis were constructed in accordance with the Silver standard (BBF RFC 23). At the time of publishing, it is the most widely adopted standard for BioBrick assembly in yeast. The assembly standards are maintained by the BioBrick Foundation, but implemented by the Registry of Standard Biological Parts (see below). Further information on the assembly standards and compatibilities can be seen on the Registry website (Rettberg 2010).

BioBricks assembly is the accepted and most widely used cloning technique within the framework of the iGEM competition, but there are many other cloning techniques. Before BioBricks, Invitrogen had a recombination-based cloning range of products called the Gateway system (Hartley, Temple et al. 2000; Sasaki, Sone et al. 2004; Alberti, Gitler et al. 2007). One of the big problems with recombination-based systems is the time-consuming incubation in a host to allow the DNA fragments to assemble. Daniel Gibson and colleagues developed a ligation free cloning technique during work on chemical synthesis of the Mycoplasma genitalium genome (Gibson, Benders et al. 2008). Gibson and colleagues ordered synthetic oligonucleotide sequences in the region of 5 – 6 kb, but this technique would also work with

natural sequences. The overlapping DNA molecules are digested using a 3'exonuclease to expose overlaps. The complementary overlaps are then annealed and the joins repaired between the sequences. A similar system using a 3'exonuclease was described in 1990 by Aslanidis and de Jong (Aslanidis and de Jong 1990). A need exists for high throughput, fast DNA assembly methods and while BioBricks is a step in the right direction in terms of part composability, faster techniques will continue to be developed beyond ligation free cloning.

1.6 Thesis Outline

This Thesis is composed of 9 Chapters and as Synthetic Biology is so different from traditional engineering disciplines, effort will be made to describe the origins of the field and some of the basic ideas, concepts and technologies. A survey of current literature is conducted on the development of the field of Synthetic Biology and the role that engineering plays in an otherwise biological discipline. This is followed by an examination of previous work on measuring promoters and a look at the evolution of peroxisome form and function in Chapter 2.

A great deal of diversity exists between labs when it comes to protocols and methods. To this end, the protocols used are presented in as complete a manner as possible in Chapter 3. Two results Chapters will be shown starting with the *E. coli* promoter characterization, the proposal of a standard promoter unit (SPU) and promoter measurement in *S. cerevisiae* in Chapter 4. Chapter 5 presents the peroxisome engineering work. Chapter 6 presents a numerical analysis of the peroxisome engineering work and Chapter 7 will discuss these results. Finally, Chapter 8 covers the conclusions and further work.

2 Review of Current Literature

The previous Chapter began at the production of the first organic molecule and continued to the revolution in DNA sequencing that led to the completing of the human genome project. Most of the major discoveries have been covered that have allowed Synthetic Biology to evolve as a discipline. Synthetic Biology can be fundamentally viewed as an engineering discipline which seeks to construct complex devices from well defined modular parts analogous to other engineering disciplines. As it is so different from traditional engineering disciplines, this review will introduce some of the basic concepts and technologies unique to Synthetic Biology.

Two major areas will be covered in this Thesis: relative promoter measurement in *E. coli* and peroxisome engineering in *S. cerevisiae*. After covering the basics of Synthetic Biology, background on previous efforts to standardize promoter measurement will be examined, followed by the body of work on which the peroxisome engineering project is based.

2.1 Basic Principles of Biological System Design

One of the key concepts in the field of Synthetic Biology is the abstraction hierarchy. The analogy most often used is that of computers and computer networks (Andrianantoandro, Basu et al. 2006). At the base level, proteins and genes are analogous to the physical layer of components on a circuit board such as resistors, transistors, diodes and LEDs. At the next level, biochemical reactions can be compared to logic gates, by which computer systems do almost all processing. Above that, biochemical pathways are like integrated circuits where the whole system has been integrated onto a chip. Cells are analogous to single computers and tissues exhibit cell-to-cell communication properties like a computer network. This analogy works well, but does not account for the way biology executes commands. The genome in a cell stores code in the same way that a hard drive in a computer does, but it also assembles the physical hardware through transcription and then translation. The connection between information storage and function execution (hardware assembly) is therefore much closer in biology. This analogy is useful in terms of visualizing the levels of complexity within a biological system, but

is not strictly true on all levels. The information contained in the genetic code has the ability to modify the physical system to a great extent; this is an emergent property of evolution. Such an analogy does not exist in computational terms as there is no feedback loop from software to hardware design within any single computer.

This abstraction hierarchy is important as each level can be examined and understood independently. Computers and other electronic devices can be built without knowing all the base physical properties of the underlying technology. Networking and systems administration are done without intimate knowledge of integrated circuit design. These analogies seek to illustrate that computers can be used without understanding how they work. Knowledge of Molecular Biology techniques and genetics at the base pair level is necessary to build even the simplest devices, limiting the people who can access biology as a technology and what they can achieve in a given time.

As there is not currently any way to build synthetic biological devices without a base level of genetics knowledge, this section will cover the basic system design principles used in this project. Any genetic construct with the purpose of expressing a protein is composed of a minimum of four elements. In its most simple form this sequence is: promoter – RBS – coding sequence – terminator as shown in Figure 2.1:

Figure 2.1: Basic genetic construct layout

2.2 Measurement in Synthetic Biology

The BioBricks standard and the Registry of Standard Biological Parts has allowed new devices and systems to be engineered and in an open-source modular format not seen in biology before. Composable parts and a functional hierarchy have allowed scientists to begin to tackle the problem of how to engineer biological systems. Databases of parts and the ability to assemble them together form the backbone of Synthetic Biology. However, accurate and useful

characteristics of these parts are often not reported. The value of the ability to accurately measure and quantify these parts cannot be understated:

> "A critical early advance at the beginning of the nineteenth century was precise measurement and analysis of compounds as they reacted. For example, by collecting and precisely weighing the carbon dioxide and water that formed upon combustion of organic molecules, it became possible to determine the atomic compositions of these molecules, and therefore their empirical formulas" (Yeh and Lim 2007).

As was previously discussed, a great deal of work building new systems and devices uses transcriptional logic. However, not all Synthetic Biology projects focus on transcriptional logic devices. Using RNA, Christina Smolke has constructed an RNA-based framework for engineering ligand-controlled gene-regulatory systems called riobzyme switches (Win and Smolke 2007). These devices can perform computation and have been configured to for the regulation of cell growth and to sense small metabolite production. Chris Voigt and his group have investigated using ribosome binding sites to engineer new genetic function (Salis, Mirsky et al. 2009). They developed a method that allowed for the predictive design of synthetic ribosome binding sites, enabling rational control over the protein expression levels in *E. coli*. Durai and colleagues are investigating the creation of zinc finger-based nucleases to enable better engineering of specific integration vectors in mammalian and plant cells by gene targeting (Durai, Mani et al. 2005). New devices cannot be built from off-the-shelf parts if there is no information on how these parts function. Two elements of these problems become important: Robust measurement of part characteristics and the standardization of these measures.

Standards are important in engineering and one of the easiest ways of transmitting information about a part or device and the standards to which it is designed is to use a datasheet. For example, the manufacturer SKF (SKF 2010) includes a datasheet with every part that allows designers and engineers to specify the exact bearing for the application. This datasheet includes all dimensions and tolerances, loading characteristics, speed ratings and fatigue load data. With this information, an engineer can specify the correct bearing for the design situation without having to resort to lab testing. The importance of defining the minimum information to characterize a part was illustrated in a news and views piece by Adam Arkin:

> 'If suitably designed, a standard can also lead to the abstraction of a composite element's behaviour into a few key functions and requirements, thereby greatly simplifying the design and analysis of the engineered system. If the abstractions are chosen just so, they may form a complete mathematical framework for design, as Boolean logic does in electronic engineering' (Arkin 2008).

As with SKF bearings, this quote highlights the importance of minimizing the information to what is necessary from an engineering standpoint. In biology, this minimum information could ideally contain data on the load the construct puts on the cell in terms of cellular resource utilization, codon utilization, Polymerase Per Second (PoPS) utilization, possible cytotoxicity, protein-protein interaction information, etc… However, there are currently no standards for all of these metrics, nor protocols to reliably measure all the listed properties. Information in biology is much more context specific than engineering. Although data sheets may eventually be important in biology, the context of the information is difficult to represent in a meaningful manner. For example: the organism, growth conditions, growth phase, single nucleotide polymorphisms (SNPS) in the gene of interest and many more parameters define the performance characteristics of the genetic element, yet in the most simple definition of a part, this contextual information is ignored.

In Synthetic Biology, the best example of part characterization is a datasheet that was created by Barry Canton for part BBa_F2620 (Canton, Labno et al. 2008). The cell-cell communication device is a composition of 5 BioBrick parts and is described as a PoPS receiver. A copy of the datasheet can be seen in Appendix 1: BBa_F2620 datasheet. This work shows an example of the parameters that need to be measured to fully characterize a BioBrick part and the utility of having the parameters recorded in a single location so that the modular device can be used to design a larger system. This datasheet represents the minimal information necessary to describe the function of a BioBrick part.

This part characterization took two graduate students approximately 12 person-months to complete (B. Canton, personal communication). The datasheet they produced is the first and possibly the only one of its kind. Resources currently do not exist to perform the same level of

work on the almost 10,000[2] parts in the Registry of Standard Biological Parts. To do this, a high throughput method of part characterization needs to be employed. But before starting high throughput assays, the standards necessary for each part and category of part need to be determined by the community. This is a 'chicken or egg' scenario, as people will not use parts without characterization information, but characterization information will not be available until the community decides what information if needs about each part.

Currently, the easiest parts to standardize are those that are self contained, modular and can either be screened using a high throughput technique or using a method that is easily reproducible across all labs. Promoters were therefore selected for standardization.

2.3 Promoter Standardization in Synthetic Biology

As previously mentioned in this Thesis, promoters have been used as tools in molecular biology since the mid 1970s when the advent of restriction enzymes made recombinant DNA work possible. Researchers discovered that changing the regulation of genes could affect their function, through over-expression and under-expression. Many promoters have since been discovered and categorized. However, there is no measure or unit, no meter, kilo or Ohm by which promoters can be compared.

The lac operon allows *E. coli* to metabolize and transport lactose across the cell membrane (Monod and Cohn 1952). It consists of a promoter, an operator, the genes lacZ, lacY, lacA and a terminator. LacY and lacA encode β-galactosidase permease, a membrane bound transport protein that pumps lactose into the cell and β-galactosidase transacetylase, an enzyme that transfers acetyl-CoA to β-galactosides. LacZ encodes the β-galactosidase enzyme, a large 464-kDa tetrameric protein.

β-galactosidase is a hydrolase enzyme that cleaves lactose into glucose and galactose. The synthetic compound o-nitrophenyl-β-D-galactoside (ONPG) is also recognized as a substrate by the enzyme. It is cleaved to yield galactose and o-nitrophenol ($HOC_6H_4NO_2$), which has a yellow color. When there is an excess of ONPG relative to the enzyme, the production of o-

[2] Although there are approximately 10,000 parts in the Registry (before the 2010 iGEM competition), many do not work and are not part of the annual parts distribution.

nitrophenol per unit time is proportional to the concentration of β-galactosidase, therefore the yellow color can be quantified to determine the level of β-galactosidase activity. If the lacZ promoter is replaced with a different promoter, its activity can be assayed by the level of yellow color. This ONPG/β-galactosidase assay was first described in a book by Jeffrey Miller (Miller 1972):

$$1\ Miller\ unit = 1000 * \frac{(Abs_{420} - (1.75 * Abs_{550}))}{(t * v * Abs_{600})}$$ **Equation 1**

Where:

- Abs_{420} is the absorbance of the yellow o-nitrophenol at 420 nm
- Abs_{550} is the scatter from cell debris, which when multiplied by 1.75 approximates the scatter observed at 420 nm
- t = reaction time in minutes
- v = volume of culture assays in ml
- Abs_{600} is the optical density of the culture at 600 nm

While this assay puts a numerical value on promoter activity, it is difficult to determine exactly what this number means. Many variables in the experimental setup can have an effect on the result of the Miller assay. Temperature, pH, the growth media, the way the promoters have been constructed and several other factors can affect the assay. Giacomini, Corlich and colleagues examined several of these experimental parameters and discovered that from a quantitative point of view, the results are markedly affected by variations these conditions (Giacomini, Corich et al. 1992). There will always be degrees of variation in all of the experimental conditions, rendering comparison between experiments, let alone labs, problematic.

In an attempt to create a better Miller assay, Eustice and Feldman used the chromogenic substrate chlorophenol red-β-D-galactopyranoside (CPRG) instead of ONPG. Their results showed the assay to be more sensitive than the standard Miller, but no effort was made to standardize the results to a common protocol (Eustice, Feldman et al. 1991).

In a review of 80 published papers on β-galactosidase assays, Serebriiskii and Golemis found the techniques could discriminate low, medium and high affinity interactions (Serebriiskii and Golemis 2000). They discovered at least six different protocols and researchers were reporting

Miller assays using either ONPG or CPRG. In many cases, the values reported in Miller units were incomparable as they were not compared against a common reference standard.

While Serebriiskii and Golemis reported problems with comparing the results of Miller assays conducted in different labs without a common reference standard, they didn't go as far as proposing one. The challenge is therefore deciding how to make an assay that will take into account or cancel out all factors external to the result of the test.

2.4 Organelle Engineering

Engineering new cells, cell free chassis and molecular pathways have been attempted in molecular biology for several years. Much of the focus of organelle creation has been focused on the self replicating aspect of the cellular compartments (Walde, Wick et al. 1994). Some groups have been looking at the creation of artificial vesicles, not from the point of view of organelle engineering, but as an analogy for primitive cell growth and division (Hanczyc and Szostak 2004). There are also groups looking to recreate the function of the Golgi apparatus *in vitro* using microfluidics, magnetic particles and recombinant enzymes (Martin, Gupta et al. 2009). Scientific American called this work the first synthetic organelle in a recent edition (Choi 2009). Another group created artificial organelles using polymers that were internalized and remained stable in macrophage cells for up to 48 hours (Ben-Haim, Broz et al. 2008). There have also been several recent reviews on artificial cells, reproducing vesicles and protocells (Roodbeen and van Hest 2009; Stano and Luisi 2010).

Several iGEM teams have also investigated building organelles. In 2009, the University of Lethbridge iGEM team aimed to create synthetic organelles in cyanobacteria by introducing micro-compartments with the purpose of creating a bio-battery (Wieden, Fischer et al. 2009). Also in 2009, the University of Toronto set out to design, model and construct a bacterial micro-organelle based system for metabolic engineering to co-localize selected enzymes (Parkinson, Cromar et al. 2009). By far the most successful organelle engineering project was from the 2007 the UCSF team. They set out to make an artificial organelle by expressing a higher eukaryote lipid phosphatase MTM in *S. cerevisiae*. They were successful in their attempts and produced fluorescence images of GFP tagged synthetic organelles in yeast (Chen 2007). However, none of the groups mentioned have looked at using peroxisomes as a base for organelle engineering.

2.5 Yeast (Saccharomyces cerevisiae) Synthetic Biology

The most common experimental organism to be used in Synthetic Biology is the gram negative bacteria: *E. coli*. This bacterium is used to assemble and copy DNA due to its high growth rate and ease of transformation. However, it is a prokaryote and cannot be used to study some aspects of biology that require more complex protein formation and folding. *S. cerevisiae*, or budding yeast, is the most well known and understood eukaryotic model organism and has been used in recombinant DNA work for some time. It is also the most used eukaryote in Synthetic Biology and as a consequence, a database of parts is being built up to support work in the area. BioBrick constructs are still assembled in *E. coli*, but yeast parts are made by taking sequences from the genome of *S. cerevisiae* and building them into BioBrick parts through PCR amplification.

Yeast was chosen as an experimental organism for this project as it has common ancestral roots with humans. There has been a considerable degree of evolutionary divergence between the two organisms, but the two genes that are used in the peroxisome project, *PEX3* and *PEX19*, have the same functions in humans and yeast. Later in the Thesis, the homology of the two protein pairs will be examined and compared

2.6 Peroxisome Engineering in S. cerevisiae

Before discussing peroxisomes, it is important to define an organelle: a specialized subunit within a cell, usually but not always surrounded by its own lipid bilayer. This membrane surrounds proteins specific to the organelle in order to allow specialized functions to proceed away from the cytoplasm. There are generally thought to be two types of organelles: autonomous and endomembrane compartment derived. The first group multiplies by growth and division and is inherited in much the same way as the nucleus. It includes mitochondria, Endoplasmic Reticulum (ER) and chloroplasts. If these organelles are lost, they cannot be produced *de novo* by the cell. The second group is derived from the ER and cannot reproduce autonomously. This group includes the Golgi apparatus, lysosomes/vacuoles, endosomes and plasma membrane. Peroxisomes are unique in that they exhibit characteristics from both groups and have not firmly been categorized as one of the other.

Peroxisomes are small, inducible organelles that can be found in almost all eukaryotic cells. They were first described as microbodies when imaged using electron microscopy in mouse kidney tissue sections (Rhodin 1954). They are in a sub-category of organelles classified as microbodies which also includes glyoxysomes and glycosomes. As they have no genome, all proteins must be imported across the membrane. The term 'peroxisome' was first used in 1966 after biochemical analysis showed that most hydrogen peroxide in the cells is degraded in peroxisomes (De Duve and Baudhuin 1966). Some of the functions of the organelle have a high degree of evolutionary conservation, lending them well to interspecies investigations (Gould, Keller et al. 1990; Tabak, Hoepfner et al. 2006). Figure 1.1 shows an electron micrograph of the yeast *Yarrowia lipolytica* containing peroxisomes.

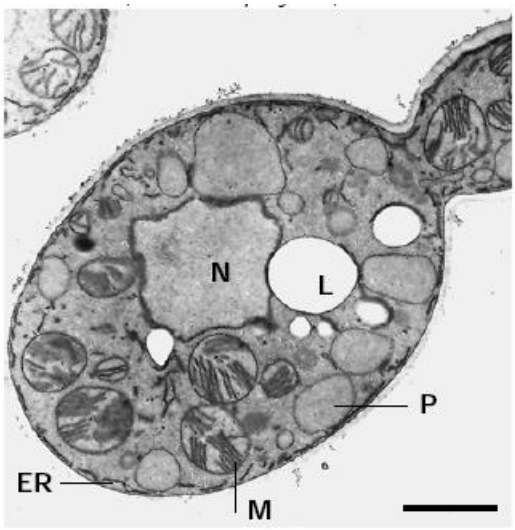


Figure 2.2: Electron micrograph of peroxisomes in *Y. Lipolytica*. Key: ER, endoplasmic reticulum, L, lipid droplet, N, nucleus, M, mitochondrion, P, peroxisome (Titorenko and Rachubinski 2001)

2.7 Peroxisome Chemistry and Cellular Function

Several essential chemical reactions take place in peroxisomes. As per their name, they contain produce, use and regulate hydrogen peroxide through various enzymes and pathways. Equation 2 shows the oxidative reaction where various enzymes use molecular oxygen to remove hydrogen from specific organic substrates (R), resulting in the production of hydrogen peroxide.

$$RH_2 + O_2 \rightarrow R + H_2O_2$$ **Equation 2**

Catalase uses the hydrogen peroxide generated from other enzymes to oxidize other substrates such as phenols, formaldehyde and alcohols. The peroxidative reaction is given in Equation 3.

$$H_2O_2 + R'H_2 \rightarrow R' + 2H_2O$$ **Equation 3**

Hydrogen peroxide is highly reactive oxygen species that is toxic to cells. It is toxic due to its ability to form reactive oxygen species (ROS) and because it is lipid soluble and can cross cell membranes. Hydrogen peroxide based oxidative stress initially results in DNA and mitochondrial damage. When cells are subject to greater oxidative stress they can repair, it can eventually lead to ATP depletion and cell death. For this reason there is a close regulation of the concentration of hydrogen peroxide. Equation 4 shows the reaction where catalase converts accumulated hydrogen peroxide to water and oxygen.

$$2H_2O_2 \rightarrow 2H_2O + O_2$$ **Equation 4**

Peroxisomes range in size from 0.1 – 1 μm and depending on cell type, can take on a spherical, tubular or cubic shape (Kurbatova, Dutova et al. 2005). Their size, number and protein content will also vary based on the metabolic need of the cell. Various biological routes of investigation have identified 32 proteins, called peroxins, which have been shown to be involved in peroxisome biogenesis (Heiland and Erdmann 2005). Using mass spectrometry, 45 peroxisomal membrane proteins have been identified in *S. cerevisiae*, suggesting that there may still be other PMPs with an unknown function (Schafer, Nau et al. 2001). When grown in the yeast media YEPD (Yeast Extract, Peptone and Dextrose/glucose) and its variants, peroxisomes are not essential organelles in *S. cerevisiae* hence they are well suited to genetic manipulation.

Peroxisomes are responsible for several important biochemical functions. They use hydrogen peroxide to carry out oxidative reactions, destroying any excess with the enzyme catalase. They also have a role in synthesizing specialized phospholipids in humans as well as degrading excess amino acids. They are discrete but contain no DNA or ribosomes. All proteins involved in their

biogenesis and maturation are imported from the cytosol (Alberts, Johnson et al. 2002). Peroxisomes also contribute to the β and alpha oxidation of fatty acids and the oxidation of bile acids and cholesterol (Hogenboom, Romeijn et al. 2002; Kovacs, Olivier et al. 2002; Kovacs and Krisans 2003; Wanders 2004). For a review on human peroxisome chemistry and disorders, see (Wanders and Waterham 2006). In humans, a lack of functional peroxisomes leads to disorders such as Zellweger syndrome, neonatal adrenoleukodystrophy, pipercolic acidemia and Refsums disease (Ghaedi, Honsho et al. 2000; Muntau, Mayerhofer et al. 2000; Weller, Gould et al. 2003).

2.8 Peroxisome Biogenesis

Since their discovery, the mechanism behind the creation of peroxisomes has been controversial. During cell growth, peroxisomes double in number and are segregated between the mother and daughter cells in a process termed 'peroxisome replication'. During normal cell growth, peroxisomes are scattered throughout the cells but when division begins, they migrate towards the bud in a Myo2 dependent manner. Peroxisomes are segregated from mother and bud along the actin cytoskeleton using the class V myosins (Myo2 and Myo4) in preparation for cell division (Hoepfner, van den Berg et al. 2001). Myo2 attaches to the peroxisome through its interaction with the protein Inp2 and carries them to the bud along actin tracks (Fagarasanu, Mast et al. 2010). In order to diffuse these organelles back into the mother cell from the bud, Myo2 is not immediately released after reaching its destination and only a few peroxisomes in the cell are transferred to the bud during division. Later, peroxisomes are transferred from Myo2 to other devices that anchor the organelles in the bud region (Fagarasanu, Fagarasanu et al. 2006). Peroxisome replication is separate to peroxisome proliferation, which is an increase in peroxisome number in response to environmental or intracellular conditions (Fagarasanu, Fagarasanu et al. 2007). Peroxisome multiplication or biogenesis describes any increase in the number of peroxisomes in a cell.

The view of the community on peroxisome biogenesis has switched many times between two competing theories. Each side can produce evidence to support their model. The first supports peroxisomes being autonomous organelles that multiply by growth and division (Kim, Mullen et al. 2006; Motley and Hettema 2007). The second states that peroxisomes bud *de novo* from the ER (Geuze, Murk et al. 2003; Kunau 2005; Tabak, Hoepfner et al. 2006).

Initially, the ER was not thought to play a role in the formation of peroxisomal proteins and peroxisomes were observed to form by division in yeast (Lazarow and Fujiki 1985). The opinion then shifted so that every membrane must originate from a pre-existing membrane (Borst 1989). The changing viewpoint often came when new techniques become available. During the 1990s, the debate continued with the addition of peroxisome deficient strains of *S. cerevisiae* that were made by deleting the *pex* genes, but it was not resolved with papers continuing to support both models.

In *S. cerevisiae*, the best data set for the *de novo* synthesis model appears in the paper: 'Contribution of the endoplasmic reticulum to peroxisome function' (Hoepfner, Schildknegt et al. 2005). Hoepfner and colleagues used real time fluorescence microscopy to track the two key peroxins involved in peroxisome biogenesis, Pex3p and Pex19p, from the ER until they formed fully functional peroxisomes. They followed the intercellular route of newly made YFP-tagged Pex3p and Pex19p using real time fluorescence microscopy in *pex3Δ* and *pex19Δ S. cerevisiae* strains. Pex3p was first observed concentrated in foci in the ER. These Pex3p focal points bud off in a Pex19p dependent manner before developing into fully functional peroxisomes. Pex19p was first enriched at the Pex3p foci on the ER and then further enriched on maturing peroxisomes. When looking at the deletion strains, without Pex19p, Pex3p was stuck in the ER and no vesicles were formed. Without Pex3p, Pex19p failed to localize to the ER membrane and no peroxisomal precompartment was formed.

Hoepfner and colleagues highlighted the key role performed by these two peroxins. In summary, Pex3p and Pex19p were shown to be essential to the import of PMPs. Pex3p was first observed on the ER where it recruits Pex19p in foci. The Pex3p-Pex19p membrane complex would then allow for the import of other PMPs (Kunau 2005). Once the membrane complexes were functional, they would enable the import of the matrix proteins (such as catalayse) to produce functional peroxisomes.

In mammalian cells, the community has still not reached a consensus as to the exact mechanism of peroxisome biogenesis. The prevailing view of the community has been that of growth and

division of existing peroxisomes (Lazarow and Fujiki 1985; Purdue and Lazarow 2001). This model categorizes peroxisomes as autonomous organelles that are not reproduced from scratch. While Pex3p and Pex19 are essential for peroxisome multiplication in yeast, Pex16p is also essential in mammalian cells (Kim, Mullen et al. 2006). *S. cerevisiae* lacks a copy of Pex16p and although the yeast Yarrowia Lipolytica has a 24% amino acid identity with human Pex16p, function is not conserved across the two species (South and Gould 1999). As the work in this Thesis will be conducted in *S. cerevisiae*, Pex16p will not be considered.

Even today, the question of peroxisome biogenesis has not been finally resolved as *S. cerevisiae* can produce new peroxisomes by both *de novo* synthesis and by peroxisome growth and division (Hettema and Motley 2009).

2.9 Peroxisome Membrane Import

As peroxisomes contain no DNA or ribosomes, all proteins must be imported across the membrane to form a functional organelle. Most proteins are synthesized on free ribosomes and imported post-translationally (Lazarow and Fujiki 1985). The matrix protein import is dependent on two peroxisome targeting signals, PTS1 and PTS2 and their cytosolic receptors, Pex5p and Pex7p respectively (Mukai, Ghaedi et al. 2002; Miyata and Fujiki 2005). These peroxisome targeting signals (PTS) correspond with the two classes of PMPs. Class I PMPs are targeted by from the cytoplasm directly to the peroxisomal membrane by the chaperone protein Pex19p at the C-terminus (Fang, Morrell et al. 2004). The class I peroxisomal membrane import process is shown in Figure 2.3. Class II PMPs are sorted to peroxisomes via the ER in a non-Pex19p dependent manner (Fang, Morrell et al. 2004; Jones, Morrell et al. 2004). The class II peroxisome import pathway has been hypothesized to use Pex16 as the anchoring site (Matsuzaki and Fujiki 2008). There are about 20 other PMPs involved in making the import complex functional.

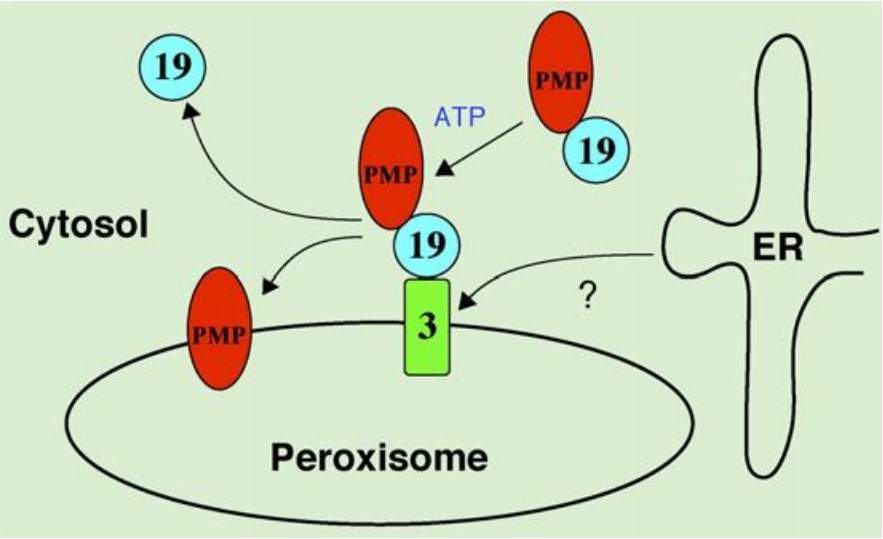


Figure 2.3: Peroxisome membrane biogenesis showing Pex19p dependent shuttling of class I PMPs to Pex3p dock. PMPs include Pex16p but exclude Pex3p in mammalian cells (Fujiki, Matsuzono et al. 2006)

Most PMPs are directed to the peroxisome by the PTS1 import tag which consists of 3 amino acids, SER-LYS-LEU, and its shuttling receptor, Pex5p. The hypothesis of shuttling receptors suggests that the import receptors Pex5p (and Pex7p) bind their cargo in the cytosol, dock to the peroxisome membrane, enter the peroxisome, release their cargo and subsequently exit the peroxisome (Marzioch, Erdmann et al. 1994). A RFP-PTS1 fusion protein will be used as a positive control to test for presence of peroxisome functionality in this Thesis.

2.10 Pex3p and Pex19p in S. cerevisiae

Pex3p (initially called PAS3) was found and recognized to be essential to peroxisome formation by Hohfeld and co-workers (Hohfeld, Veenhuis et al. 1991). This discovery was made by isolation of *PEX3* mutants and the cloning of the *PEX3* gene by functional complementation.

As with most of the results in the peroxisome community, there are two camps with regard to the function of Pex3p. The first suggests that the docking domain of Pex19p binds to Pex3p (Fang, Morrell et al. 2004; Hoepfner, Schildknegt et al. 2005). The second supports Pex3p being targeted to the peroxisome independent of Pex19p (Heiland and Erdmann 2005; Fujiki, Matsuzono et al. 2006). This Thesis is based on the work of Fang, Morrell and co-workers and Hoepfner, Schildknegt and colleagues with the assumption that Pex19p has a docking domain

for Pex3p. They also state that Pex19p mediates the transport of cytosolic Pex3p to peroxisomal/ER bound Pex3p via Pex19p.

Pex19p was also discovered to be essential to peroxisome formation along with its human analog (Gotte, Girzalsky et al. 1998). Gotte and colleagues tested the human analog (gene product HK33) and although an initial variant of the protein didn't rescue peroxisomes in yeast, a chimeric version was successful in rescuing peroxisome function, suggesting the possibility of human-yeast peroxisome interaction.

Pex19p interacts with all class I PMPs through a consensus motif that is found at least once in each membrane peroxin (Sacksteder, Jones et al. 2000; Rottensteiner, Kramer et al. 2004). Pex19p has two key domains that are of interest in this project. It has a docking domain for Pex3p in the 56 amino acids proceeding the NH_2-terminal. It also has a binding domain for class I PMPs from amino acids 57 – 299 (Fang, Morrell et al. 2004). It functions as a targeting sequence receptor for peroxisomal membrane proteins.

Peroxisome formation from a pre-peroxisome compartment requires Pex19p in *S. cerevisiae* (Hoepfner, Schildknegt et al. 2005). Without Pex19p, Pex3p can not leave the ER (Tam, Fagarasanu et al. 2005). However, the nature of this Pex19p mediated Pex3p release is currently unknown.

2.11 Pex3p and Pex19p in Mammalian Cells

In mammalian cells, the interactions of Pex3p and Pex19p are slightly different to yeast. Mammalian cells have an extra essential peroxisomal protein, Pex16p, whose function and mechanism of action do not appear to be completely known. The difference between yeast and humans may lie in the existence of a specialized ER subdomain that is not present in yeast cells (Geuze, Murk et al. 2003). The lack of this subdomain may have an effect on the formation of human peroxisomes in yeast.

Pex3p is anchored in the peroxisome membrane within a 33 amino acid N-terminal sequence. This region is sufficient to target Pex3p to the peroxisomes (Soukupova, Sprenger et al. 1999). Interactions with Pex19p are mediated by the cytosolic protein binding domain (Fujiki,

Matsuzono et al. 2006). Pex3p is imported into the peroxisome in a Pex19p independent manner, making it separate to the class I *PSTI* import system. This finding is supported by the observation that in mammalian cells, a transient inhibition of Pex19p has no effect on the import of Pex3p into pre-existing peroxisomes (Jones, Morrell et al. 2004).

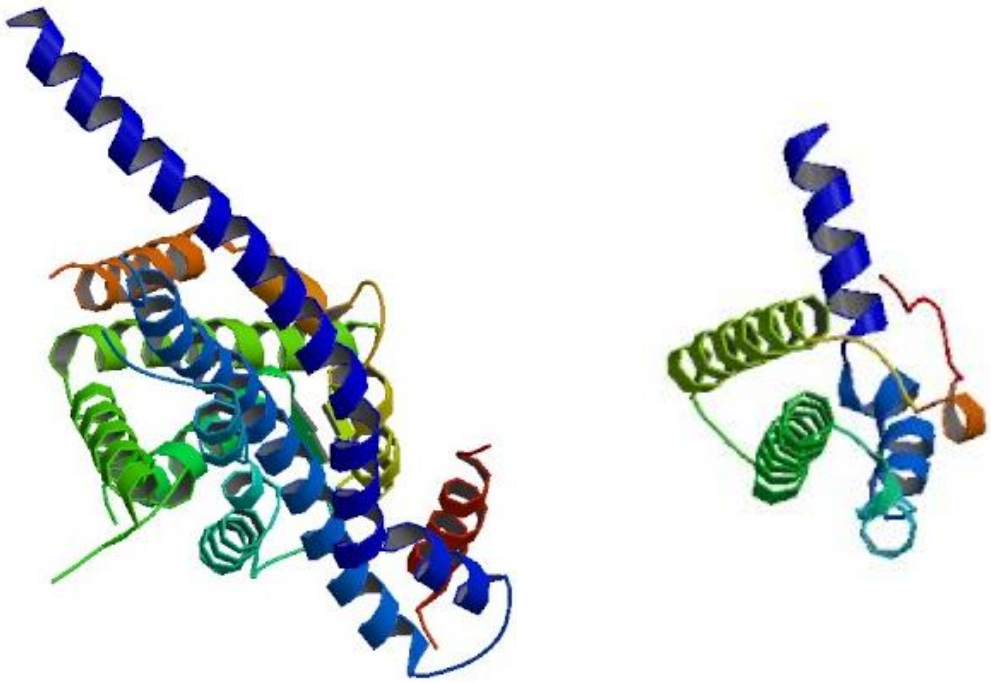
Pex19p is a protein that shows a broad binding specificity for peroxisomal membrane proteins and is hydrophilic in nature (Gotte, Girzalsky et al. 1998; Sacksteder, Jones et al. 2000). Pex19p is mostly found in the cytosol, yet a small but significant proportion is also to be found on the peroxisomal membrane (Matsuzono, Kinoshita et al. 1999). Pex19p contains a farnesylation consensus motif at the C-terminus that provides structural integrity to the protein, but principally serves to target the protein to the peroxisome (Banerjee, Kessler et al. 2005). Pex19p is currently thought to have several functions.

Fransen and colleagues investigated the interactions of Pex3p and Pex19p in mammalian cells in a paper entitled: 'Analysis of Human Pex19p's Domain Structure by Pentapeptide Scanning Mutagenesis' (Fransen, Vastiau et al. 2005). In summary, they used transposon mutagenesis to build a library of human Pex19 alleles that coded for random in-frame pentapeptide insertions. Their work revealed the presence of a tripartite domain structure in Pex19p with an amino terminal domain that binds to Pex3p, essential for docking to the peroxisome membrane. Pex19p was also shown to contain a central domain that competes with Pex5p and Pex13p for Pex14p binding and a carboxy-terminal domain that interacts with multiple PMPs. They were unable to determine if the protein functioned as a chaperone or not. Finally, they discovered that Pex19p contains two distinct binding sites for Pex3p, indicating that it may bind PMPs in several places for several purposes.

2.12 Human Pex3p and Pex19p Protein Structures

Human Pex3p and Pex19p have only recently had their structures resolved by X-ray diffraction. Figure 2.4 A shows the resolved structure of the protein complex where the group have reported that PEX3 adopts a novel fold best described as 'a large helical bundle' (Schmidt, Treiber et al. 2010). Figure 2.4 B shows the structure of the folded C-terminal part of the receptor reveals a globular domain that displays a bundle of three anti-parallel long helices (Schueller, Holton et al.

2010). With the resolution of these structures now complete, the possibly of re-engineering the protein binding domains is now open.



A) **B)**

Figure 2.4: A) Pex3p in complex with Pex19p, B) Pex19p. The structure of Pex3p in complex with Pex19p was resolved using X-ray diffraction (Schmidt, Treiber et al. 2010). Pex19p was also resolved using X-ray diffraction (Schueller, Holton et al. 2010)

2.13 Peroxisome Engineering Summary

This section has summarized the function of peroxisomes and highlighted the difference between yeast and mammalian peroxisome function. Peroxisome morphology and chemistry has been discussed. The targeting of peroxisomes with the class I PMP system has been described. The essential peroxisome biogenesis proteins in humans and yeast, Pex3p and Pex19p, have been described.

Human Pex19p has previously been shown to bind to yeast Pex3p (Fransen, Wylin et al. 2001; Fransen, Brees et al. 2002). A chimeric yeast Pex19p has been shown to work with human Pex3p (HK33) (Gotte, Girzalsky et al. 1998). However, the focus of the community has been on

the investigation of the function of Pex3p and Pex19p in multiple organisms, sometimes using mutant versions of the proteins. Interactions of wild-type recombinant human Pex3p with wild-type Pex19p have been reported using pull-down assays and surface plasmon resonance analysis. Wild type and two recombinant mutants of Pex3p showed K_D values of 3.4 nm, 1080 nm and 66.2 nm respectively (Sato, Shibata et al. 2008). Information of the kinetic binding affinities for *S. cerevisiae* Pex3p and Pex19p has not been reported in the literature.

The interest in the project lies in the binding domains for Pex3p on Pex19p and vice versa. With the interspecies binding specificities for Pex3p and Pex19p being lower than the species specific binding affinities, there is a possibility of observing bimodality in the biogenesis of peroxisomes. So far in the literature, putting both proteins from yeast and mammalian cells in the same organism with the aim of observing the resulting organelle formation has not been attempted. This Thesis fits into the peroxisome community by building on accumulated knowledge of the peroxins Pex3p and Pex19p. This project has combined fluorophore tagged versions of yeast and human Pex3p and untagged yeast and human Pex19p proteins in an attempt to build artificial peroxisomes.

Finally, Bimodality and bistability are interesting in this context as engineering dual functionality can produce interesting effects and emergent behaviors. The first example of engineered bistability is Gardner's toggle switch in *E. coli* (Gardner, Cantor et al. 2000). The introduction of two modes of operation for their switch allows for the introduction of elementary DNA processing in the cell, without the difficulty of having to engineer protein-protein interactions. This type of switch can then be used to build networks with more complex functions (Ellis, Wang et al. 2009). In the context of this project, bimodality would allow for the creation of two populations of peroxisomes. If these populations could be selectively targeted, novel biochemistry could be created and targeted while maintaining endogenous peroxisomes. Each population of peroxisomes could allow for different chemical environments to be created, depending on the desired reactions. Bimodality is therefore important in biology as a tool to engineer novel functions and interactions.

2.14 Yeast and Peroxisomes Image Analysis Techniques

Image analysis and automated image analysis is necessary in this Thesis to identify yeast cells and the peroxisomes they contain. Scripts will be written to automate the process where possible in an attempt to remove the human element from quantitative processing. Previous work has examined the development of image processing techniques for yeast cell morphology (Ohtani, Saka et al. 2004). They developed a script that detects the cells from the background using an edge detection and thresholding system. Ohtani and colleagues focused on measuring parameters associated with yeast morphology to categorize cells into cells cycle phases. Beyond yeast morphology, systems have been designed to automatically identify subcellular features and organelles based on pattern recognition from fluorescence images of mammalian cells (Murphy, Boland et al. 2000).

Image analysis has been applied to peroxisomes in the past. Techniques have been described for the quantitative evaluation of peroxisome morphology in rat livers, using a combination of automated image analysis and immunocytochemistry (Beier and Fahimi 1987; Beier 1992; Beier and Fahimi 1992). This system is designed to work principally with tissue samples when investigating changes to the peroxisome enzyme proteins. Automated scripts have also been described to determine the number of peroxisomes per cell in *S. cerevisiae* using a composition of out-of-focus brightfield images and single channel fluorescence images of GFP peroxisomes. A script was written to identify peroxisomes from a fluorescent image and select them within a region from a brightfield image (Niemisto, Selinummi et al. 2006). This technique was interesting as it used an algorithm to separate the mother from the bud during processing, but their identification of the peroxisomes was poor in general. Incorrect thresholding and a lack of sufficient filtering led to non-peroxisome areas in the GFP images being falsely tagged as peroxisomes.

2.15 Aims and Objectives

This Chapter has examined the basic concepts of the field of Synthetic Biology, from an engineering point of view. The background literature for promoter measurement in *E. coli* and organelle engineering in *S. cerevisiae* has been covered. This Thesis will aim to investigate the

development of two foundational technologies in Synthetic Biology: standardized promoter measurement and the creation of an artificial organelle.

This aim will be accomplished through the following objectives:

- Measure *E. coli* promoters and compare results to those of other labs using an *in vivo* reference standard;
- Measure a collection of *S. cerevisiae* promoters by applying techniques from the *E. coli* promoter measurement work;
- Develop and build a set yeast strains capable of expressing human and yeast peroxisomes;
- Image these yeast strains;
- Validate the nature of the peroxisome population through numerical analysis;
- Determine if artificial peroxisomes have been created.

3 Materials and Methods

3.1 Growth Conditions and Media

Luria-Bertani (LB) (Sambrook, Fritsch et al. 1989) was prepared using 10 g of bacto-tryptone, 5 g of yeast extract and 10 g of NaCL in a total volume of 1 l of double distilled water. YEPD medium contained 2% dextrose (Fluka, ≥99% molecular biology grade), 1 % yeast extract (Fisher Scientific, Molecular genetics grade) and 2% bacto-peptone (Oxoid, molecular biology grade). YPDA medium contained 2% dextrose, 1% yeast extract, 2% bacto-peptone and 0.1% adenine (Sigma, minimum 99%). SD medium contained 2% dextrose, 0.67% yeast nitrogen base without amino acids (Fluka) and 0.2 % synthetic complete amino acid powder, without YNB (US biological). YNO medium contained 0.1% oleic acid (VWR, extra pure grade), 0.5% Tween 40, 0.1% yeast extract and 0.67% yeast nitrogen base without amino acids. YPGO medium contained 0.1% oleic acid, 0.5% Tween 40, 3% glycerol (Fischer Scientific, analytical reagent grade) 1% yeast extract and 2% yeast nitrogen base without amino acids. STYO medium contained 0.1% Tergitol NP-40 () 0.005% yeast extract, 0.67% yeast nitrogen base without amino acids, 0.2% synthetic complete amino acid powder without YNB and 0.1% adenine. S-KAc medium contained 0.67% yeast nitrogen base, 0.2 % synthetic complete amino acid powder, without YNB, 0.2% potassium acetate (Fischer Scientific, analytical reagent grade).

E. coli strains were grown from – 80 ^{o}C stocks onto solid media agarose LB plates. For experiments, strains were transferred from solid media to liquid LB cultures and grown at 37 ^{o}C for 12 – 1 hours at 200 rpm. Unless otherwise stated, *S. cerevisiae* strains were grown from – 80 ^{o}C stocks for 16 hours at 200 rpm and 30 ^{o}C. Antibiotics were prepared, filter sterilized, stored at – 20 ^{o}C and used at predefined concentrations (Table 3.1) unless otherwise stated. Yeast and bacterial stocks were stored long term at – 80 ^{o}C in 30% v/v glycerol solution (yeast) and 20 % v/v glycerol solution (bacteria).

3.2 Antibiotics used in this study

Antibiotic	Final concentration (µg ml^{-1})		Solvent
	E. coli	*S. cerevisiae*	
Ampicillin	50	-	Water
Kanamycin	50	-	Water
Genetecin	-	200	Water
Phleomycin	-	25	Water

Table 3.1: Antibiotics and working concentrations used in the study

Antibiotics were prepared, filter sterilized and stored at – 20 ^{o}C. All antibiotics were purchased from Sigma.

3.3 Molecular techniques

3.3.1 Plasmid purification

5 ml overnight stationary phase *E. coli* cultures grown in LB medium were harvested by centrifugation using a microfuge (Beckman and Coulter) at 13 000 x g for 5 minutes. The plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and the above microfuge. Plasmid DNA was eluted using 30 µl of the elution buffer (EB) contained in the kit. Once eluted, the sample DNA concentration was measured using a UV visualization Nanodrop (Thermo Scientific). Samples were then adjusted to a concentration of 100 ng/µl using EB. The plasmid DNA was stored at -20 ^{o}C.

3.3.2 Transformation of DNA into Bacterial Cells

Plasmid DNA was transformed into Top 10 or DH5α cells (both Invitrogen) using the enclosed protocols. DNA was incubated with cells on ice for 10 minutes and then heat shocked at 42 ^{o}C for 30 seconds then placed back on ice for 5 minutes. The cells were then incubated with SOC media (2% w/v tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) at 37 oC for 1 hour with shaking before being plated onto LB plates with the appropriate antibiotic selection.

3.3.3 Restriction Digests

Restriction digests were performed during BioBrick part assembly when two parts and during colony PCR to check for insertion success. The *EcoRI*, *XbaI*, *Spe1*, *PstI*, *BsteII* and *NotI* Fastdigest enzymes were used (Fermentas). Reactions were performed according the manufacturer's specifications using 1 µl of each restriction enzyme (1 unit), 1 µl of 10x Fastdigest buffer (Fermentas), up to 1 µg of plasmid DNA and the reaction volume was completed to 20 µl using ddH_2O. The reactions were incubated at 37 ^oC for 5 minutes, run on an agarose gel to verify fragment size and purified to remove remaining enzymes.

3.3.4 Agarose Gel Electrophoresis

Purified DNA was analyzed using 0.8 – 1.2 % (w/v) agarose gels and ethidium bromide (0.5 µg ml^{-1}) in 1 x TAE buffer (per litre at 50 x concentration: 242 g tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA adjusted to 1 l using ddH_2O). Eight parts DNA were typically added to one part loading buffer (30% v/v glycerol, 0.25% w/v bromophenol blue and 0.25% w/v xylene cyanol) prior to loading the gel. A 1 kb or 2 log DNA ladder (both New England Biolabs) were used according to the manufacturers specifications to size DNA products. Gels were run at V until sufficient migration had occurred, determined by the position of the leading dye front. Gels were visualized using a UV transilluminator and Chemidoc gel documentation system (Bio-Rad). When required, DNA was excised from gels and purified using a QIAquick Gel Extraction Kit (Qiagen) and purified using a microfuge according to the manufacturer's instructions.

3.3.5 Lithium Acetate Transformation (LiOAc) of S. cerevisiae Cells

Cells were transformed using a variant of the Ito and colleagues LiOAc protocol (Ito, Fukuda et al. 1983). Cells were grown in 10 ml YEPD cultures overnight to a density of 1-2 $x10^7$ cells/ml. They were harvested by centrifugation at 4000 x g for 4 minutes. Cells were washed in 10 ml of ddH_20 and harvested at 4000 x g for 5 minutes. After resuspension in 1 ml of ddH_20, the cells were harvested in a microfuge at 13000 x g for 10 seconds. The cells were washed with 1 ml of LiOAc/TE (100 mM Tris-HCl, 10 mM EDTA, 0.1 M LiOAc), harvested by centrifugation at 13000 x g for 10 seconds and ressuspended in 50 µl of LiOAc/TE. Salmon sperm DNA (10 mg/ml) was prepared by boiling for 4 minutes at 100 ^oC. 2 µl of Salmon sperm was added to 4 µl of the requisite plasmid at a concentration of 250 ng/ µl and 300 µl of 50% w/v PEG 3350

g/mol (Sigma, molecular biology grade) in LiOAc. The mixture was vortexed for 60 seconds and incubated at 30 ^{o}C on a roller shaker for 30 minutes. After incubation, 35 µl of DMSO (Sigma, molecular biology grade, 99.9%) was added and the tubes were vortexed for 60 seconds. After a 15 minute heat shock at 42 ^{o}C, the cells were harvested by centrifugation at 2000 x g for 5 minutes and the supernatant removed using a vacuum aspirator. The cells were resuspended in 100 µl of ddH_2O and plated on selective plates and incubated overnight at 30 ^{o}C.

3.4 LoxP Flippase Removal Protocol

Phleomycin is a glycopeptide antibiotic from the bleomycin family. To remove a gene using the LoxP protocol, page-purified primers with 40 nt base-pair regions of overlap with the region of were designed and ordered (Invitrogen). Knockout cassettes were constructed using PCR amplification of the selected EUROSCARF vector (pUG66 – ampicillin resistance in *E. coli*, phleomycin resistance in yeast) with the Pex3 and Pex19 knockout primers shown in Table 3.2. Cells were transformed with the PCR knockout cassette using the previously described LiOAc transformation protocol onto auxotrophic selective medium plates. A single colony was selected and grown in 5 ml of liquid YEPD medium to mid-log phase (1-2 x 10^7 cells/ml). The cells were transformed with the flippase plasmid, pPS-3098, using the previously described LiOAc protocol and grown on phleomycin selective medium plates for 48 hours. A single colony was selected and inoculated into 1 ml of YEP-GAL, incubated a 30 ^{o}C for 2 hours and added to 9 ml of YEDP and incubated overnight. 10 µl of this culture was plated onto YEPD plates and incubated overnight at 30 ^{o}C. From this plate, 20 colonies were selected and incubated on YEDP plates overnight at 30 ^{o}C. Using the same velvet, the YEDP plates were replica plated onto the original auxotrophic selective medium. Colonies that grow on the YEDP but not the auxotrophic plates have had the gene in question successfully deleted and have lost the phleomycin resistance cassette. This protocol can be repeated several times to produce a strain with multiple knockouts without using selectable markers.

3.5 Pex3 and Pex19 knockout primers

Primer function	Sequence
Pex3 forward	ATGGCCCCAAATCAAAGATCACGTTCGCTTCTGCAGAGAC**cagctgaagcttcgtacgctgcagg**
Pex3 reverse	AGGCTTGAAGGAAAACGAGCTGGAGACGCCAAAGTTGCTG**gcatagggccactagtggatctgatatcacc**
Pex19 forward	ATGCCAAACATACAACACGAAGTAATGAATGAAAACGAGT**cagctgaagcttcgtacgctgcagg**
Pex19 reverse	TTGTTGTTTGCAACCGTCGGTTAATTCCTTATCAAGGTTC**gcataggccactagtggatctgatatcacc**

Table 3.2: Pex3 and Pex19 Knockout primer sequences. The capital sequence shows the 40 nt region of homology with the chromosome integration site. The lower case bold sequences represent a 25 nt homology with the EUROSCARF cassettes to allow for PCR construction of knockout cassettes.

3.6 BioBrick Standard Assembly

BioBrick parts were received on a plasmid in backbone vectors pSB1A2, pSB1A3 or pSB3K3. When BioBrick parts were assembled, the insert was cut from the backbone by performing a restriction digest as explained previously using either restriction enzyme pairs *EcoRI* and *SpeI* or *XbaI* and *PstI* (Fermentas). The vector was cut using the complement restriction enzyme pairs *EcoRI* and *XbaI* or *SpeI* and *PstI* (Fermentas). The insert reaction was comprised of 17.0 µl of plasmid vector DNA (500 ng/ µl or 8.5 µg), 2.0 µl of 10 x FastDigest buffer (Fermentas) and 0.5 µl (0.5 units) of both restriction enzymes (Fermentas). The vector digest reaction contained 14.5 µl of ddH_20, 20 µl of vector (500 ng/ µl or 10 µg of DNA), 4.0 µl of 10 x FastDigest buffer, 0.5 µl of calf intestinal alkaline phosphatase (NEB) and 0.5 µl (0.5 units) of both restriction enzymes (Fermentas).

Once digested, the vector was purified using a QIAprep Spin Miniprep Kit (Qiagen). The DNA concentration was measured using a Nanodrop (Thermo-Scientific) and normalized to 100 ng/µl. The insert was run on a gel and after the correct sized fragment excised, purified using a QIAquick Gel Extraction Kit (Qiagen) and again measured using the Nanodrop. The excision and purification can lower the concentration of DNA to below 20 ng/µl so it is important to start with a high concentration. After the insert and vector were restriction digested and purified, they

were ligated together using a Rapid DNA Ligation Kit (Roche). Once transformed and selected, the *E. coli* strains were stored long term in 20% w/v glycerol at -80 ^{o}C.

3.7 GFP Paraformaldehyde Fixation

Cells were cultured according to previously described protocols and media. Paraformaldehyde fixation was performed by centrifuging cells at 13,000 x g for 30 seconds and removing the supernatant using a vacuum aspirator. 100 µl of diluent paraformaldehyde mix (10% v/v 37% paraformaldehyde (Sigma, molecular biology grade), 0.143 M NaCl (Fisher, analytical reagent grade), adjust to pH 7.5 using KOH), 1.43 M EDTA (Sigma, molecular biology grade 99+%), 71.41 mM HEPES (Fischer, 99+%)) was added and the cells were vortexed for 30 seconds. Cultures were incubated at room temp for 2-5 minutes followed by centrifugation at 13,000 x g for 30 seconds and removal of the supernatant using a vacuum aspirator. Cells were washed once in 0.5 – 1 ml of KPO_4/sorbitol (21.8% w/v sorbitol (Fisher, 98+%), 1.4% w/v K_2HPO_4 (Fisher, 99%), .23% w/v KH_2PO_4 (Fisher, 99%)). This cell sorbitol mix was resuspend in 30-50 µl of KPO_4/sorbitol and stored for up to four weeks. Prior to microscopy, the cells were resuspended by vortexing for 30 seconds. Paraformaldehyde fixation can reduce fluorescence signal, but in the case of peroxisomes in live yeast cells, fixation resulted in a greater signal as the peroxisomes could be immobilized in the cells and longer acquisitions used without loss of resolution.

3.8 Mounting Media Protocol for All Strains

Strains grown in oleate and surfactant based media were suspended for microscopy in mounting media of same composition without amino acids, glycerol or yeast extract. YNO mounting medium contained 0.1% oleic acid, and 0.5% Tween 40. YPGO mounting medium contained 0.1% oleic acid and 0.5% Tween 40. STYO mounting medium contained 0.1% Tergitol NP-40 and 0.1% oleic acid. For strains that had been grown in sugar based media, the KPO_4/sorbitol from 3.7 was used to mount cells.

4 Measuring Promoter Activity Using an *In vivo* Reference Standard in *E. coli* and *S. cerevisiae*

The foundations of modern engineering, since the industrial revolution, have been built on standardization, decoupling and abstraction. Systems biology has evolved further with standardized programming languages for models such as CellML and SBML, but implementing standards *in silico* will generally remain easier than *in vivo*. The first widespread standard in Synthetic Biology is for the functional compositional of DNA parts: BioBricks™. These parts are available to order in a catalog from the Registry of Standard Biological Parts at MIT. However, this standard only relates to DNA part assembly. There are few standards yet in Synthetic Biology relating to the quantification of DNA function.

4.1 Contributions to this Chapter

This Chapter discusses the creation of a standard measure for promoter activity in *E. coli*. It is based on the author's contribution to the paper: Measuring the activity of BioBrick promoters using an *in vivo* reference standard (Kelly, Rubin et al. 2009). Jason R Kelly built the *E. coli* strains, established the theory behind the relative measurement system. The other authors on the paper (Adam J Rubin, Joseph H Davis, Caroline M Ajo-Franklin, John Cumbers, Michael J Czar, Aaron L Glieberman, Dileep D Monie and Drew Endy) participated in the multi-institution experiments. In the yeast promoter measurements section, Jake Wintermute constructed the promoter constructs and transformed them into the yeast strains. The author had assistance with the flow cytometry experiments from Bruno Afonso. All data analysis in the yeast promoter experiments was conducted by the author.

4.2 E. coli Promoter Measurement Introduction

Promoter expression levels have been previously measured in *E. coli* and experiments to describe the strength of these promoters have been created (Chapter 2.3). One problem with using Miller units is that there no precise way of measuring promoter activity with repeatable accuracy; different days, under different conditions or between labs (Giacomini, Corich et al.

1992). There is a great degree of variation in methodologies and reagents used to measure promoter expression in Miller units and the technique is sensitive to these variations. This section describes a method to measure the activity of BioBrick promoters using an *in vivo* reference standard.

As previously stated, there are many promoters in the Registry of Standard Biological Parts and the number grows every year. Some of these promoters have been characterized but no single characterization protocol has been followed to allow their performance to be compared. A reference standard was created called the Relative Promoter Unit (RPU) and a basic measurement protocol to make measurements under standardized conditions was suggested.

4.3 Measurement Kit Background

A major concern with measuring any biological variable is the effects of the environment on that variable. The effect of environmental parameters such as growth rates, agitation, temperature, humidity and oxygenation levels need to be considered when concerning fluorophores. Maturation rates of green fluorescent protein molecules are affected by all of the previously listed factors, so controlling or normalizing is necessary to ensure consistent fluorescence.

4.3.1 Assumptions

One way to address the environmental variables is to test the promoter activity using a relative standard for each measurement. Most of the environmental variables can be eliminated as the samples will be grown under the same conditions using the same equipment. One area where the samples may differ is if a 96 well plate is used and sealed using a plastic film that does not allow oxygen into the well. This may result in a difference in oxygenation levels across the wells in the plate, so a breathable membrane was chosen. Alternately, experiments could be conducted using a small volume of mineral oil in each well on top of the growth media and cells to reduce evaporation in the wells and allow oxygenation of the samples.

Along with the environmental variables, promoter elements need to be standardized to make validating assumptions. The reference promoter (BBa_J23101) was originally derived from a promoter collection constructed by the 2006 Berkeley iGEM team. They constructed a family of constitutive promoters by isolated sequences from a small combinatorial library. They then

measured the promoter activity using relative RFP fluorescence measurements. The reference promoter was assembled to GFP and the measured promoter expressed in the same background strain on the same plasmid (in different wells) so that the GFP maturation rate can be assumed to be constant. Both promoters are also on the same backbone plasmid so the plasmid copy number is also constant. Both promoters have been standardized to have the same transcription initiation sites and identical downstream sequences so the rate of mRNA degradation may be assumed identical. A final assumption is that the immature GFP is stable so that protein degradation is negligible compared to dilution due to cell growth. The half-life of GFP has previously been reported at approximately 26 hours (Corish and Tyler-Smith 1999).

The above assumptions allow comparisons to be made between promoters based solely on the difference between promoter activity levels. This is the basis for the comparative promoter measurement experiments in *E. coli*. A further numerical explanation can found with derived equations in Appendix 4. In any biological system, a degree of noise will affect behavior of individuals and the population as a whole. Noise can be broken down into two main categories intrinsic and extrinsic. Intrinsic noise creates a difference between two reporters in the same cell and extrinsic is where two reporters are affected equally in the same cell, but differences are observed from cell to cell in a population (Swain, Elowitz et al. 2002). In the case of this system, sources of noise exist, but the results reported by Kelly are none the less comparable between experiments conducted on different days and even using different media conditions (changes in carbon sources).

4.3.2 Strains and Promoters

In order to use a relative measurement, several elements in the testing protocol were standardized. Invitrogen TOP10 was selected as the *E. coli* strain for the project as it is easy to work with, freely available and in common usage among iGEM teams. A selection of constitutive promoters from the Registry of Standard Biological parts was chosen.

Strain	Construct name	Promoter
TOP10	-	-
TOP10	BBa_I20259	BBa_J23113
TOP10	BBa_I20269	BBa_J23150
TOP10	BBa_I20270	BBa_J23151
TOP10	BBa_I20268	BBa_J23102
TOP10	BBa_I20260	BBa_J23101

Table 4.1: Promoters and constructs used in the multi-institution experiments. Note BBa_J23101 is the reference promoter.

All promoters came from the J. Chris Anderson promoter collection that was developed by the 2007 UCSF iGEM team. The collection is from a family of constitutive parts that were isolated from a combinatorial library. Five promoters and the TOP10 control were chosen from this collection for the experiment: BBa_ J23113, BBa_J23150, BBa_J23151, BBa_J23102 and BBa_J23101 (the reference promoter). They were chosen to represent a good cross-section of expression levels and they had already performed some basic characterization to assess activity levels. The experimental strains and a TOP10 control as shown in Table 4.1 were distributed by Jason Kelly to the institutions taking measurements in an agar stab. The backbone plasmid for the strain collection is shown in Figure 4.1.

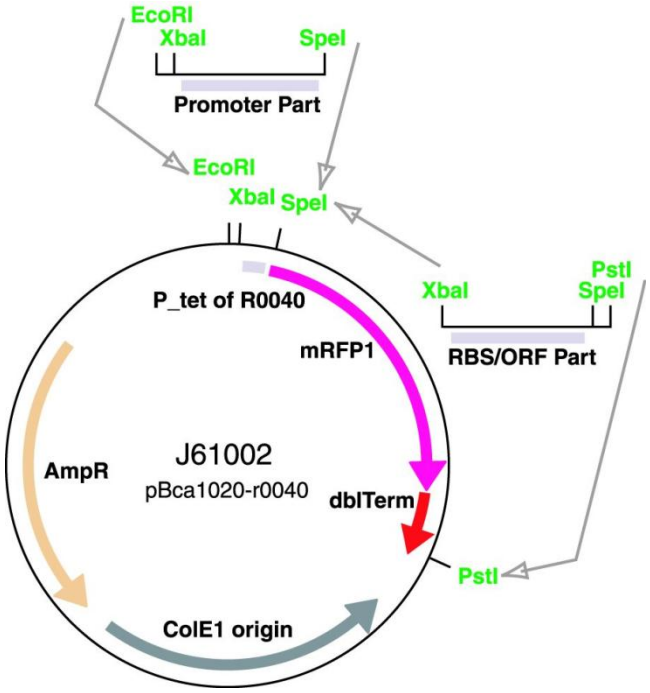


Figure 4.1: Plasmid map showing backbone used in collection construction (Anderson 2006).

4.3.3 BD LSR-II Flow Cytometry

Measurements were taken for the multi-institution experiments using flow cytometry. A BD LSR-II with a 488 nm laser was used for the experiments in the case of this Chapter. No attempt was made to standardize flow cytometry equipment across institutions. Seven parameters were recorded: forward scatter height, area, side scatter height, area, GFP height, area and time. Forward and side scatter were used to calibrate the detector voltages. These voltages were adjusted to the correct particle size, which in this case is *E. coli*. The particle size is a function of the detector voltage, so adjustments to the voltage are made by measuring the minimum number of detection events, (10 000) and calibrating the voltage relative to the forward vs. side scatter plots. The GFP area data were used for promoter analysis.

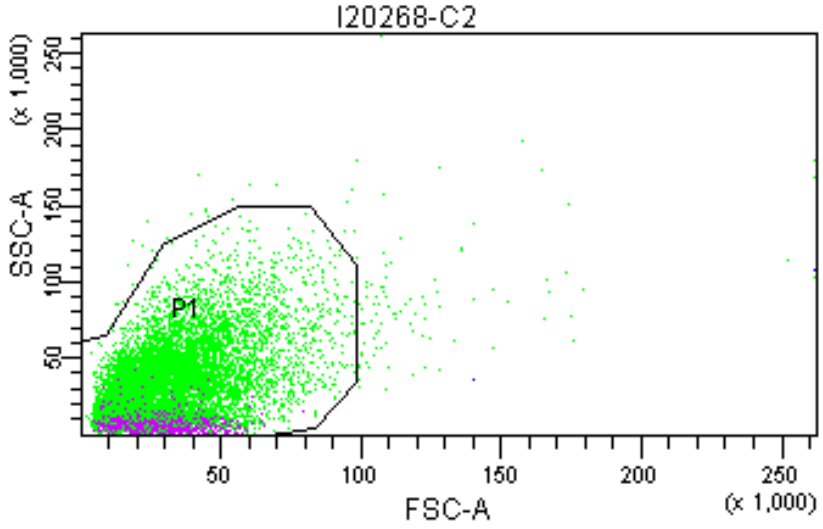


Figure 4.2: Example gating used during flow cytometry analysis of *E. coli* promoter collection. This gating was created based on the TOP10 background strain, the BBa_I20260 reference promoter

Three colonies for each promoter had been grown and prepared as described in the methods section earlier in this Chapter. They were used to calculate the level of promoter activity based on the relative difference between the TOP10 base strain and the BBa_I20260 reference promoter. A gating was applied to the samples based on the forward and side scatter channels and the negative control. An example of this gating can be seen in Figure 4.2. Finally, the mean of fluorescence intensity per cell was calculated for each culture.

4.3.4 Flow Cytometry Data Analysis

The author participated in the development of the multi-institution measurement protocol and acquired data for publication, but was not involved in the analysis. This section will briefly describe the analysis for completeness. Three colonies were grown and measured in the flow cytometer for each promoter. The average GFP area measurement was calculated using the flow cytometry software and this value input into an Excel spreadsheet. The mean and standard deviation were calculated for each promoter using the average from each well.

4.4 Multi-Institution Measurement Results

The reference promoter was measured using a variety of carbon sources in the media, on different days and using a plate reader and a flow cytometer. The only data shown here will be the multi-institution results.

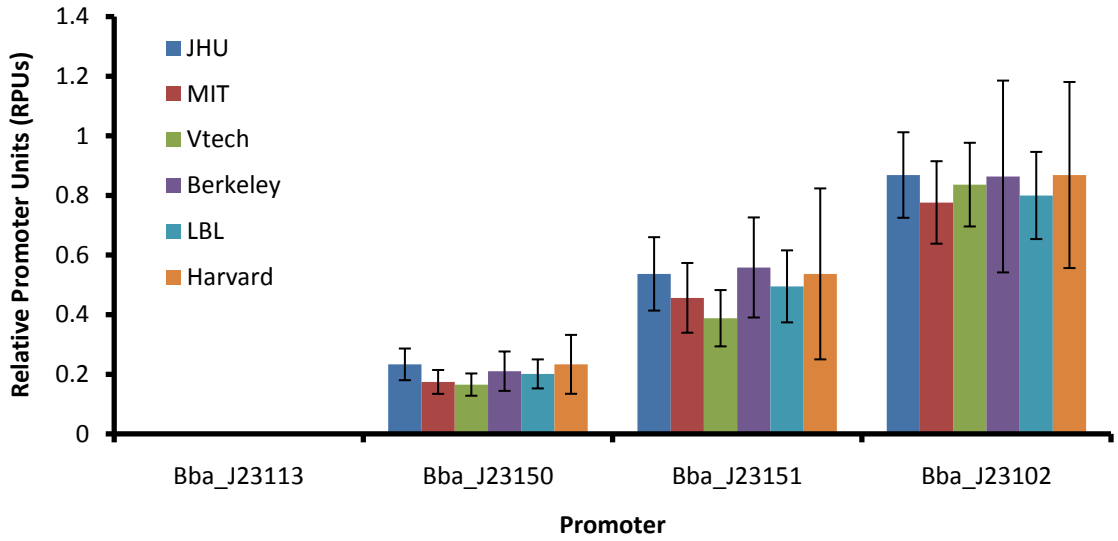


Figure 4.3: Multi-institution flow cytometer results where the error bars represent the 95% CI of three replicates. Note: the authors' results are displayed under "Harvard"

The institutions involved in the survey were: Johns Hopkins University (JHU), Massachusetts Institute of Technology (MIT), Virginia Institute of Technology (Vtech), UC Berkeley, Lawrence Berkeley National Laboratory (LBL) and Harvard. Six strains containing five promoters and the TOP10 control were measured. The error bars are larger than expected due to the use of only three replicates. BBa_23113 is shown as zero because the promoter is not strong enough to be differentiated from the control. This represents one of the limits of the technique, promoters can only be measured if they can detected by flow cytometry when GFP is being expressed. The reference promoter is used to normalize the samples (to 1 RPU) so it is also not shown. There is less than a two-fold range across the other three promoters measured by the labs (BBa_J23150: 0.14 – 0.23; BBa_J23150: 0.38 – 0.606; BBa_J23103: 0.77 – 0.96) (Kelly, Rubin et al. 2009).

4.5 E. coli Promoter Measurement Kit

After the reference standard was created, it needed to be publicized and distributed to the community. History is littered with examples of multiple standards being introduced simultaneously, from the creation of wide gauge railways in the 19th Century through to the competing formats of Blue-ray and HD-DVDs. In order to create a standard that would be accepted and maintained, it was decided to follow the example of the standard Ohm resistor that was introduced in the 1850's to solve the problem of finding breaks in the transatlantic telegraph cable. The Ohm kit contained a one Ohm resistor made of solid brass and all the necessary instructions on how to use the kit. This kit was so successful in introducing a standard to a wide audience, that it was deemed the best model to follow for a promoter measurement kit. This kit was created containing all the instructions and the physical DNA spotted onto chromatography paper to allow users to measure their own promoters relative to the reference standard. The reference kit was distributed to all the teams in time for the 2008 iGEM competition. A description of the standard reference kit can be found on pages 191 and 196. Feedback was obtained on the kit from the Harvard team. They transformed the test constructs from the measurement kit into DH5α *E. coli* cells and obtained similar values to the test promoter activity levels when measured using a fluorescence plate reader.

4.6 Measuring the Activity of Promoters in S. cerevisiae

In the previous sections, a technique was devised to measure the strength of promoters in *E. coli*. When this work was completed, it was sought to apply these techniques to other situations. The technique was adapted to a collection of yeast promoter strains and the results were collected using a similar measurement protocol. The GFP collection was made by constructing an insertion cassette containing a GFP fusion and a selectable marker (Huh, Falvo et al. 2003). The GFP was fused to the gene by using the PCR constructed homologous recombination cassette. The promoter strength can be approximately measured by the level of the fluorescence, but this is also a measure of the fused protein. This approximation was the reason why a collection of strains was constructed to more accurately test the strength of various *S. cerevisiae* promoters.

4.7 Introduction to Promoter Measurement in S. cerevisiae

After developing a technique to measure promoter activity in *E. coli*, investigations were conducted to apply similar techniques to a collection of yeast strains constructed to measure the

promoter expression levels. This strain collection is constructed in a different way to the *E. coli* project as each strain contains two fluorophores, RFP and YFP. The aim was to compare the expression level of a constitutive *ADH1* driving RFP to the promoter of interest driving YFP.

The experiments in this section represent a preliminary investigation into the possibility of measuring yeast promoter expression levels with the aim of creating a reference standard. Data is presented to illustrate the technique and the strains but; further experimentation is necessary before a protocol giving complete characterization of the promoter is achieved.

4.7.1 S. cerevisiae Promoter Strain Construction

The background strain for this project is 580a or FY23 and is known as B1 in the context of this project. It has the following genotype: URA3Δ52, TRP1Δ63, LEU2Δ1, GAL+. Each strain has an *ADH1* promoter with a human Kozak sequence expressing a fusion of two mCherry RFP proteins. Each strain (other than the RFP control) also contained a fusion of two mVenus fluorescent proteins expressed by a combination of either human or yeast Kozak sequence and the promoter of interest.

4.8 Yeast Promoter Collection Methods

The growth protocols were copied as closely as possible to the *E. coli* experiments, so that the same assumptions could be made when showing the relative promoter strengths. Changes were made so the technique could be performed using the high throughput sampler on the flow cytometer. Supplemental protocols can be found in the appendices on page 167.

4.8.1 96 well format work

The strain collection was assembled in triplicate in three 96 well plates. Deep well plates were inoculated with 500 µl of YEPD media and shaken at 1000 rpm at 30 ^{0}C overnight in an Appropriate Technical Resources (ATR) Multitron 2, with a 3 mm orbit. When the cultures were grown to saturation, 500 µl of 50% w/v glycerol was added to each well and after shaking, the plates were frozen at – 80 ^{o}C for long term storage. Three plates were used in case any single well or plate became contaminated or in case nine replicates were needed. A V&P Scientific Inc. 384/96 well floating head pin tool was used to inoculate the plates. The tool was

configured for 96 well operation. An automated machine technologies (AMT) MicroPlate plate filling machine was used to correctly dose 96 well plates with the required media volumes.

4.8.2 Flow Cytometry (FACS)

A BD Biosciences LSR-II with a High Throughput Sampler module (HTS) was used to acquire the FACS data. Two Lasers were used:

1. 633 nm JDS Uniphase HeNe laser (Red).
2. 488 nm solid-state laser (Green).

The forward and side scatter channels were used to calibrate the settings on the flow cytometer using cells from the 580a control wells. The detector voltage was calibrated using the *ADH1* promoter for RFP and the *THD3* promoter for YFP. The number of detection events on the flow cytometer was set to 10 000 and once all the settings were correctly adjusted, the automated high throughput sampler measured all wells in sequence.

4.8.3 Media conditions

The promoter collection experiment was initially grown in YPDA for the overnight step in the protocol. For the re-inoculation step the following day, the growth medium was changed to either synthetic dextrose or synthetic potassium acetate medium. The media were made using synthetic amino acid compositions, yeast nitrogen base and either glucose or potassium acetate as the carbon source. Synthetic media were chosen to minimize background signal and avoid unintended autofluorescence on the YFP channel that can be introduced with standard yeast media (YEDP/YPDA). The media compositions can be found in Chapter 3, page 37.

4.9 Strain details

The promoter collection was designed to test a variety of promoters and Kozak sequences, which define the translation initiation codon in eukaryotes. Each strain contains the control *ADH1* promoter, human Kozak sequence and a fusion protein consisting of two mCherry RFPs. The *ADH1* promoter is known to be constitutive in *S. cerevisiae* during growth on glucose, although not when glucose in the media becomes depleted (Denis, Ferguson et al. 1983). It controls expression of the alcohol dehydrogenase 1 enzyme in yeast. Acetaldehyde is reduced to ethanol by *ADH1* to regenerate NAD^+ allowing continuation of the glycolytic cycle. The collection was

split into two major groups with each promoter driving constructs that had either the human or yeast Kozak. Within these two groups, there are a variety of promoters.

Strain name	Promoter	Kozak	Fluorophore	Strain name	Promoter	Kozak	Fluorophore
B1	-	-	-	B2	-	-	-
Y1	p*ADH1*	*hs*Kozak	2xRFP.m				
Y2	p*ADH1*	*hs*Kozak	2xYFP.v	Y18	p*ADH1*	*sc*Kozak	2xYFP.v
Y4	pGAL1	*hs*Kozak	2xYFP.v	Y20	pTDH3	*sc*Kozak	2xYFP.v
Y5	pTDH3	*hs*Kozak	2xYFP.v	Y21	pTEF1	*sc*Kozak	2xYFP.v
Y6	pTEF1	*hs*Kozak	2xYFP.v	Y22	pPGK1	*sc*Kozak	2xYFP.v
Y7	pPGK1	*hs*Kozak	2xYFP.v	Y23	pMDH3	*sc*Kozak	2xYFP.v
Y8	pMDH3	*hs*Kozak	2xYFP.v	Y24	pPEX3	*sc*Kozak	2xYFP.v
Y10	pPEX8	*hs*Kozak	2xYFP.v	Y25	pPEX8	*sc*Kozak	2xYFP.v
Y11	pPEX11	*hs*Kozak	2xYFP.v	Y26	pPEX11	*sc*Kozak	2xYFP.v
Y12	pPEX19	*hs*Kozak	2xYFP.v	Y27	pPEX19	*sc*Kozak	2xYFP.v
Y13	pPEX25	*hs*Kozak	2xYFP.v	Y28	pPEX25	*sc*Kozak	2xYFP.v
Y14	pHIS3	*hs*Kozak	2xYFP.v	Y29	pHIS3	*sc*Kozak	2xYFP.v
Y15	pCTA1	*hs*Kozak	2xYFP.v	Y30	pCTA1	*sc*Kozak	2xYFP.v
Y16	pCUP1	*hs*Kozak	2xYFP.v	Y31	pCUP1	*sc*Kozak	2xYFP.v

Table 4.2: Promoter strain collection details. This table shows the YFP promoters. All strains except the controls B1 and B2 contain p*ADH1*-RFP. Note the left side of the table contains strains with the human Kozak sequence and the right, the yeast Kozak sequence.

Table 4.2 shows a graphical representation of the strain collection. The *pGAL1* strain is only present with the human Kozak sequence and the pPEX3, with the yeast Kozak sequence. As much of the work in the rest of this Thesis involves peroxisomes, there are four peroxisomal promoters: *PEX8*, *PEX11*, *PEX19* and *PEX25*. The remaining promoters were chosen as they had previously been assembled into BioBrick form.

4.10 Yeast Promoter Collection Analysis

Once data were acquired using the flow cytometer, the first stage in the analysis was to setup the correct gating. Software from the manufacturer of the flow cytometer called BD FACS-Diva was used to analyze the FCS 3.0 files. The P3 region in Figure 4.4 shows an example of the gating used to select the correct population of yeast cells. This gating was selected based on the forward and side scatter data encompassing 95% of the cells for three representative strains from the collection: *ADH1*-RFP control, p*TDH3* and p*GAL1*. Once the gating had been setup, the

RFP and YFP mean and standard deviation was generated in an automated manner for each test well on the 96-well plate. These data will be used in the analysis to show the difference between promoter expression levels.

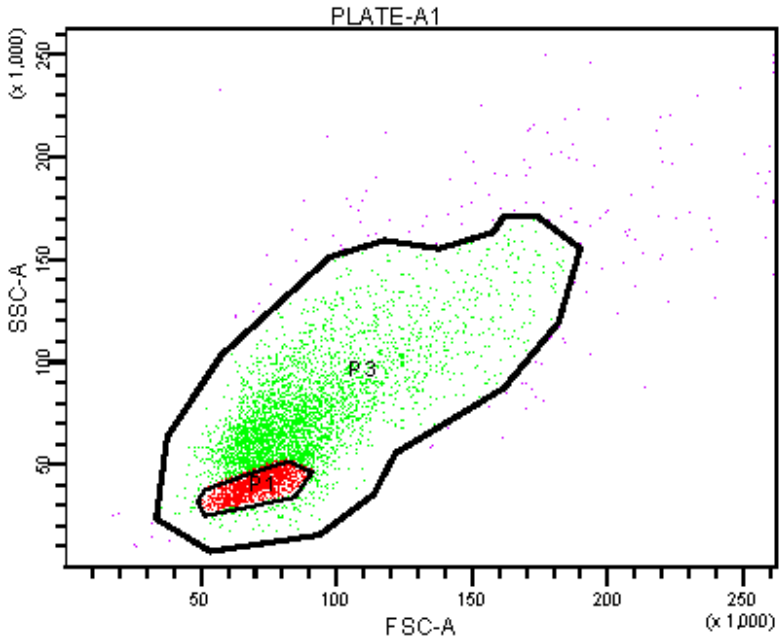


Figure 4.4: Example gating showing data from the control p*ADH1*-RFP strain. This gating was created to capture most cells in three strains: the p*ADH1*-RFP control, p*TDH3* and p*GAL1*

4.11 Characterization of the transcriptional activity of the yeast promoters in the ADH1-RFP collection

The results of this experiment are presented in two sections. The first describes the *ADH1* promoter – 2x RFP for each strain on a media to media basis and the second, the SD results for the promoter – 2xYFP strains. All constructs were integrated into the yeast chromosome. All RFP constructs were integrated into the *TRP1* loci while the YFP constructs were integrated into the *LEU1* loci using the Sikorski vectors (Sikorski and Hieter 1989). The *ADH1* – 2X RFP experiment was repeated several times with SD media to fine tune the protocols and equipment (data not shown). When the procedure was performed and verified, it was conducted using the previously described media.

4.11.1 ADH1 Promoter - RFP Control Results

As previously described, all strains have a constitutive *ADH1* promoter expressing a double RFP fusion protein. This section describes the RFP strain to strain variation and assesses using the RFP to normalize the collection. All strains in this section have the same construct: *AHD1* promoter - 2x RFP. The promoter names refer to the promoter driving the YFP protein, but in this section are only used to differentiate the strains. Figure 4.5 and Figure 4.6 show two groups of promoters, roughly dividing the plot in half. The first half contain YFP strains with the yeast Kozak sequence and the second half, the human Kozak sequence. Finally, the last three columns represent the *ADH1* promoter - 2x RFP with no YFP construct, the background strain and the media with no yeast cells. A scaling factor was set based on the SD medium average to allow comparisons with the S-YNB samples. The units of the Y-axis are shown in this relative scale. The YFP promoter results were also been adjusted using the same scaling factor. The strain collection was grown in two media conditions to assess the effect of changing the medium conditions on the promoter expression levels. S-KAc data were normalized to the SD average. Error bars in the plots show the range.

4.11.2 Constitutive ADH1 Promoter – 2x RFP Average in SD Media

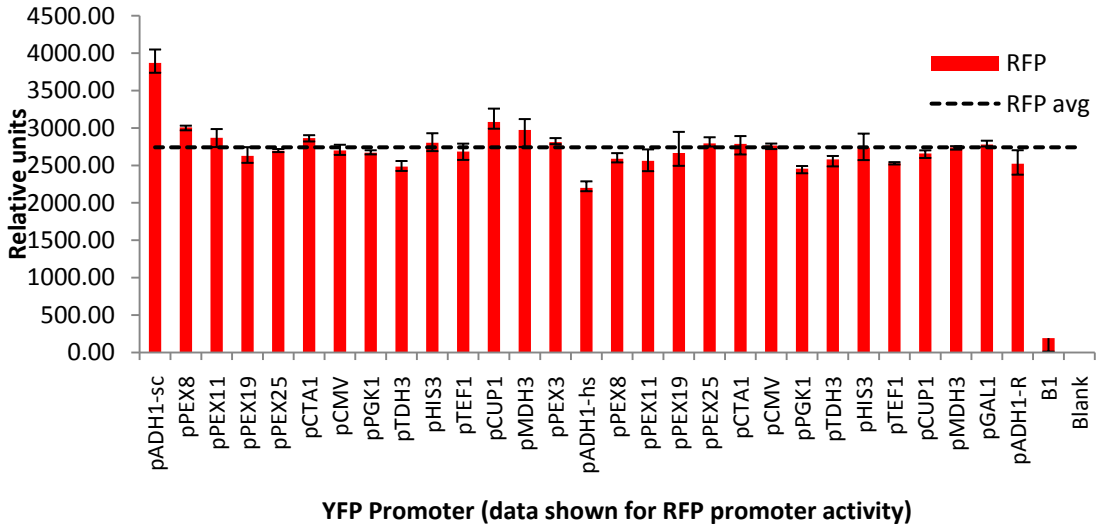


Figure 4.5: *S. cerevisiae ADH1* promoter strain collection expressing RFP (mCherry). The strains were grown in YEDP overnight and induced in SD medium for 6 hours. The collection was grown suspended in TE buffer and placed on ice until measured by the BD LSR-II flow

cytometry machine. The error bars show the range of three replicates for each strain and dashed line, the average of the collection.

Figure 4.5 shows the strain collection grown in SD medium. There is little deviation from the population average and only a small range between the three replicates. The first promoter, the yeast *ADH1* is clearly different to the rest of the population. This was likely due to a double integration of the construct or a form of fluorescent contamination. This promoter was excluded for the numerical analysis.

Media:	SD
Min	2200.33
Max	3081.67
Range (Min – Max)	881.33
Average	2570.83
Residuals squared:	2.21E+06
Error – residuals squared:	2.81E+05
Error + residuals squared:	3.69E+05
Scaling factor	1.00

Table 4.3: SD medium numerical parameter summary table.

Table 4.3 shows the numerical properties of the promoter collection in SD. The range for the SD collection was smallest of all media conditions due to the *ADH1* promoter being tightly regulated by glucose. The residuals squared were calculated from the square of the difference of the strain from the collection average. The error – is the square of the difference between the lowest replicate and the average. Error + is the same for the highest replicate. The scaling factor for SD is 1.00 as it is used as the datum to compare data with the potassium acetate medium.

4.11.3 Constitutive ADH1 Promoter – 2x RFP Average in S-KAc Media

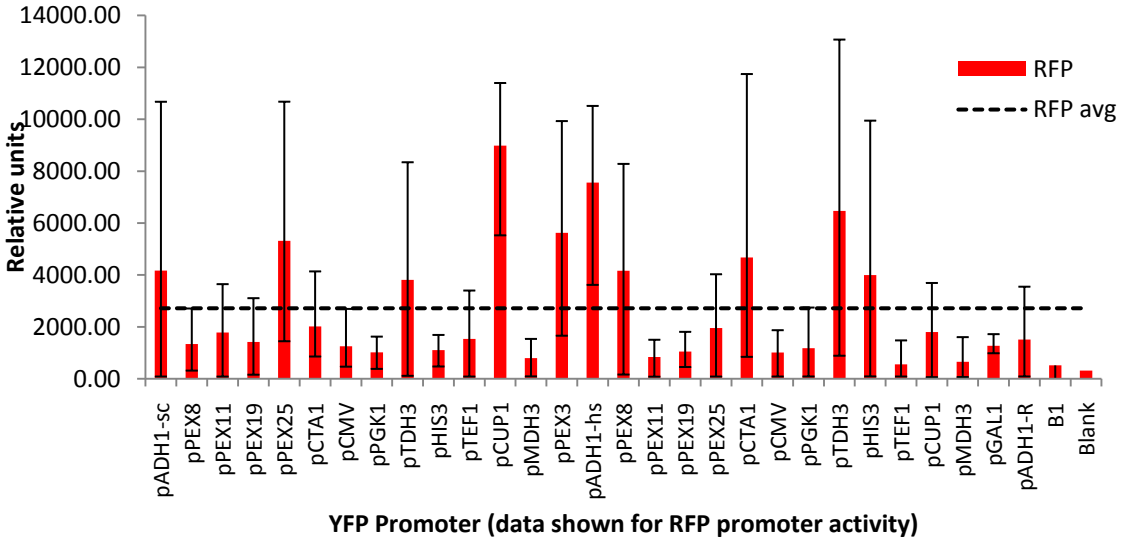


Figure 4.6: *S. cerevisiae ADH1* promoter strain collection expressing RFP (mCherry). The strains were grown in YEDP overnight and induced in S-KAc medium for 6 hours. The collection was grown suspended in TE buffer and placed on ice until measured by the BD LSR-II flow cytometry machine. The error bars show the range of three replicates for each strain and dashed line, the average of the collection.

Figure 4.6 shows the *AHD1* promoter – double RFP fusion protein average in S-KAc media. There is a clear difference between the transcriptional behavior shown by the average of the relative units in this plot when compared to SD medium in Figure 4.5. The strains exhibits a much greater range between replicates when grown in potassium acetate medium. The error indicates that the carbon source does not have a consistent regulatory effect on the *ADH1* promoter in each replicate. *ADH1* is induced by glucose, but mRNA levels have been shown to decrease when grown on a non-fermentable carbon source (Denis, Ferguson et al. 1983). This repression is not very tight when compared to induction on glucose, as demonstrated by the large an inconsistent range for each sample in Figure 4.6.

Media:	SKAC
Min	553.92
Max	8986.01
Range (Min – Max)	8432.09
Average	2570.83
Residuals squared:	1.46E+08
Error – residuals squared:	1.88E+08
Error + residuals squared:	3.09E+08
Scaling factor	1.13

Table 4.4: S-KAc medium numerical parameter summary table.

Table 4.4 shows numerical data for yeast strains grown in S-KAc medium expressing RFP under control of the *ADH1* promoter. In this case, the minimum, maximum and range are very different to those seen with SD medium. The residuals are also two orders of magnitude higher.

4.11.4 Constitutive ADH1 Promoter Analysis

Figure 4.5 to Figure 4.6 shows the normalized constitutively expressed *ADH1*-RFP in the promoter collection. Table 4.5 shows a summary of the numerical parameters that we calculated based on *ADH1*-RFP expression data in different media conditions. The minimum and maximum represent the weakest and strongest RFP signal of all strains and the range is the difference between the two. The residuals were calculated by squaring the difference between the RFP value and the RFP population average. The +/- residuals were calculated by squaring the error.

Media:	**SD**	**S-KAC**
Min	2200.33	553.92
Max	3081.67	8986.01
Range (Min – Max)	881.33	8432.09
Average	2570.83	2570.83
Residuals squared:	2.21E+06	1.46E+08
Error – residuals squared:	2.81E+05	1.88E+08
Error + residuals squared:	3.69E+05	3.09E+08
Scaling factor	1.00	1.13

Table 4.5: p*ADH1*-RFP summary table.

Strains grown in SD medium have residual values that are two orders of magnitude lower than S-KAc. This section presented results acquired by measuring promoter collection in two media conditions. When grown in glucose, the *ADH1* promoter is tightly regulated. In potassium acetate media, this regulation is almost non-existent. The aim of this experiment was to use the *ADH1* promoter as a reference to normalize the other promoters. The data in Table 4.5 show that assumption can be applied in the case of SD media, but not with potassium acetate media. Based on the above analysis of *ADH1*, the definition of 'constitutive' should be updated to encompass media conditions. Observations were made that indicate the difference between *ADH1* regulation on glucose media and on potassium acetate media. Although there is a clear indication that the *ADH1* promoter behaves differently when grown on a fermentative carbon source compared to a non-fermentative carbon source, stochasticity is likely to play a role in the regulation of this promoter.

4.12 YFP Promoter collection results

While RFP data was collected for the ADH1 promoter, the YFP channel was used to acquire data on the different promoters in the collection. YFP data from the promoter collection activity is shown for SD medium.

4.12.1 Promoter, Human and Yeast Kozak Sequence Comparison in SD Medium

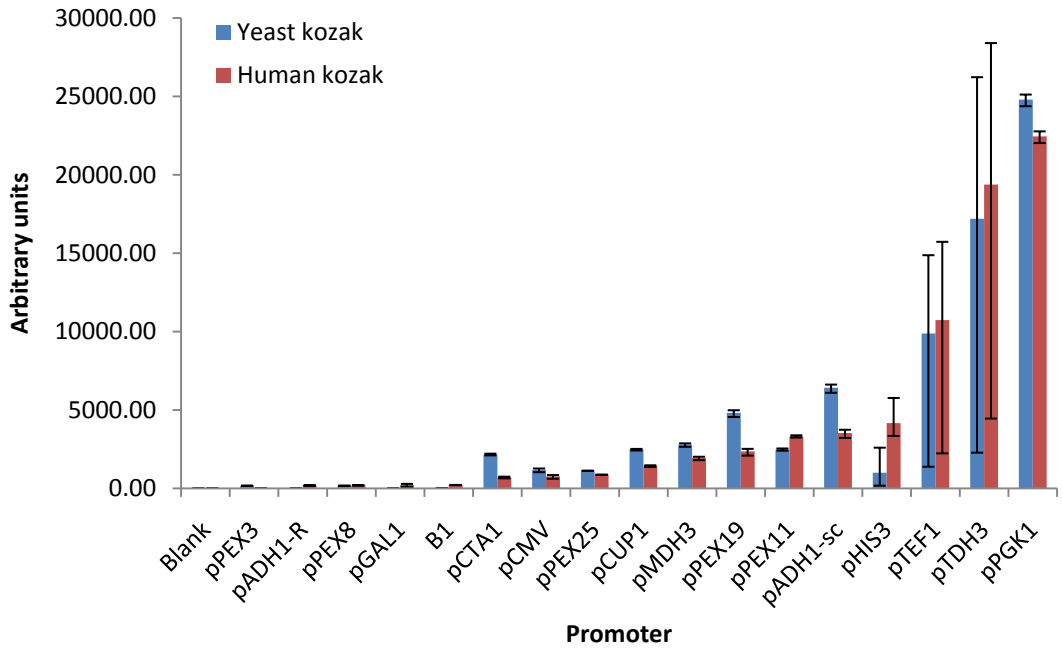


Figure 4.7: Human and yeast Kozak sequence comparison in SD media. Promoters ordered from human weakest to strongest. *S. cerevisiae* promoter strain collection expressing YFP (mVenus). The strains were grown in YEDP overnight and induced in SD medium for 6 hours. The collection was grown suspended in TE buffer and placed on ice until measured by the BD LSR-II flow cytometry machine. The error bars show the range of three replicates for each strain and dashed line, the average of the collection.

Figure 4.7 shows the results of the promoter collection in SD ordered by the relative strength of the yeast Kozak sequence. The promoters show approximately the same expression levels with each Kozak sequence with the exception of the *HIS3* promoter. As explained previously, error bars display the range of the three replicates while the solid bars represent the average. The *TEF1* and *TDH3* promoters are among the strongest, but show a great variability in the replicates. The *PGK1* promoter is the strongest with a small variation in the replicates. The other promoters are not induced at a very high level relative to the three highest.

4.13 Promoter Measurement Discussion

A method to measure the promoter activity in *E. coli* across multiple labs using a standardized reference promoter was presented. Construct BBa_I20260 (promoter only part: BBa_J23101)

was chosen as the constitutive *E. coli* reference promoter. Results from the multi institution experiments were shown, indicating that using a relative standard and measurement protocol can produce close results across labs with non-standard equipment. A measurement kit was developed and distributed to the iGEM teams during the 2008 competition and discussions with teams who used the kit reported that it worked well.

The aims of the yeast promoter experiments were to assess the promoter activity of a strain collection in different media compositions by applying knowledge gained from the *E. coli* measurement experiments and to investigate a yeast reference promoter. This work provided an indication that the *ADH1* promoter could be used as a reference promoter. However, several other promoters such as the *HXT1*, *HXT2*, *HIS3* and potentially PGK1 should also be investigated. The protocol used to test these promoters should specify glucose as the carbon source in the media as the experiments using multiple media conditions were ultimately not successful. Further work is required to be conducted to measure the promoter collection strains with single constructs. The YFP measurements were ultimately not reported in conditions other than SD. Using the Kelly method with only the YFP fluorophores could result in a more robust, reliable comparative measure for yeast promoters if a yeast measurement kit in the form of BioBrick parts could be implemented and distributed with the iGEM part collection.

5 Peroxisome Engineering in *S. cerevisiae*

Two terms will be defined and referenced throughout this Thesis. "Image-set" refers to a series of images taken under identical conditions for the purpose of data analysis and the generation of color composite images. "Data-set" refers to the quantified information produced from the "image-set" after image processing. Two imaging setups were used with different imaging characteristics. The initial work was conducted on a Nikon microscope at Harvard and will be referred to as the Nikon microscope/setup/system. The second setup was a DeltaVision microscope at the Centre Optical Instrumentation Laboratory (COIL) facility and will be referred to as the DeltaVision microscope/setup/system. These microscopy setups are described in greater detail later in the Thesis.

5.1 Contributions to this Chapter

The individual BioBrick parts involved in this project were all constructed by Jake Wintermute and Caroline Ajo-Franklin. Simple part combinations, such as RBS-promoter and gene-terminator assemblies were also constructed by Jake Wintermute. The author assembled all BioBrick constructs from simple part combinations into yeast transformation-ready vectors. The author performed all yeast transformations into YPH500α while Keiji Nishida transformed all constructs into BY4741a. All microscopy in the Chapter was performed by the author.

5.2 Peroxisome Project Background

Peroxisome biogenesis and organelle engineering was covered in detail in sections 2.4 to 2.13 in Chapter 2. This section gives a brief recapitulation of the project theory and motivation. The current theory on how peroxisomes are formed is that Pex3p is expressed at a minimum level when the cell is growing under normal (YPDA/YEPD glucose carbon source medium) conditions and is then transported by Pex19p to the ER. The Pex19p shuttle protein has a domain-recognition sequence for the membrane-bound Pex3p and a transport domain for other peroxins (including Pex3p). Pex3p targets to discrete ER-localized punctae, forming an ER subcompartment en route to the peroxisome (Hoepfner, Schildknegt et al. 2005; Tam,

Fagarasanu et al. 2005). This feedback system allows for the transport of cytosolic Pex3p to ER bound Pex3p, increasing the local concentration of the protein. One assumption for this project is that the human Pex3p can bind to the docking domain of the yeast Pex19p, but with a lower binding affinity. The other assumption is that the human Pex19p also docks to the yeast Pex3p with a lower binding affinity.

The combination of these characteristics led to the possibility of a system that could potentially demonstrate bimodal behavior. In this case, bimodal is defined as the creation of two populations of peroxisomes, where one is entirely (or mostly) human and the second, yeast. The "human" peroxisomes would not be entirely human as all peroxins other than Pex3p and Pex19p would be the yeast variants. The motivation is hence to create a new organelle that would allow the compartmentalization of harmful or useful chemical reactions away from the cytosol while maintaining a population of the endogenous organelles. Figure 5.1 shows a cartoon of the two possibilities where in A) yellow shows human peroxisomes, blue yeast and in B) hybrid human yeast peroxisomes.

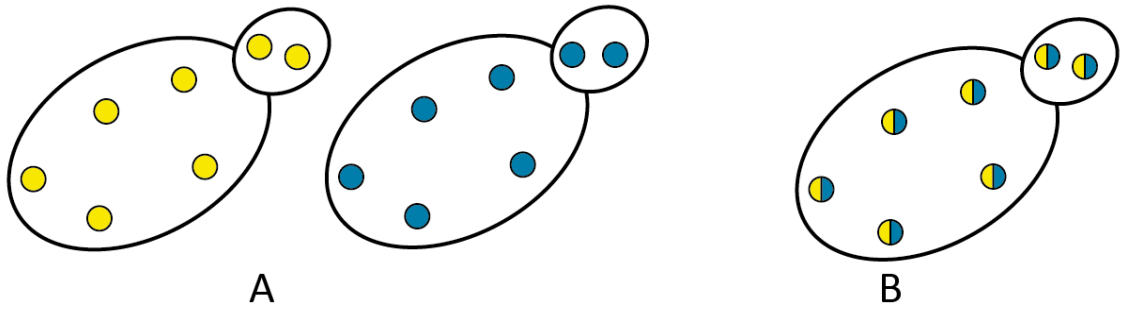


Figure 5.1: A) illustrates the outcome of the principal experiment if bimodality is observed while B) represents a homogenous population of peroxisomes.

5.2.1 Pex3p and Pex19p sequence analysis

Human Pex3p and Pex19p are known to work with each other and the binding affinities of these two proteins have been discussed in Chapter 2. Yeast Pex3p and Pex19p are also known to interact, but binding affinities have yet to be measured and reported in the literature. The two protein pairs (yeast Pex3p with human Pex19p and human Pex19p with yeast Pex19p) are known to bind to each other, but the binding affinities have also yet to be measured and reported in the literature. Bimolecular fluorescence complementation (BiFC) could be used as an

indication of the binding strength (Kerppola 2006). The Pex3p-Pex19p interaction would be well suited to characterization using this method as the human interaction strength could be verified and compared with reported values in the literature, before moving on to the yeast-yeast and human-yeast interactions. The sequences used in the project were aligned using AlignX and then analyzed for sequence homology.

Sequence analysis of the protein coding regions of human and yeast Pex3p revealed a 40.3% identity. Figure 5.2 shows a plot of the sequence homology mapped using AlignX (Invitrogen Vector NTI). The identity was mapped with a 5 amino acid moving average showing the difference in the height of the peaks. The absolute complexity is the pairwise alignment score between the two sequences. Human and yeast Pex19p were also analyzed and showed a 43.6% identity (data not shown).

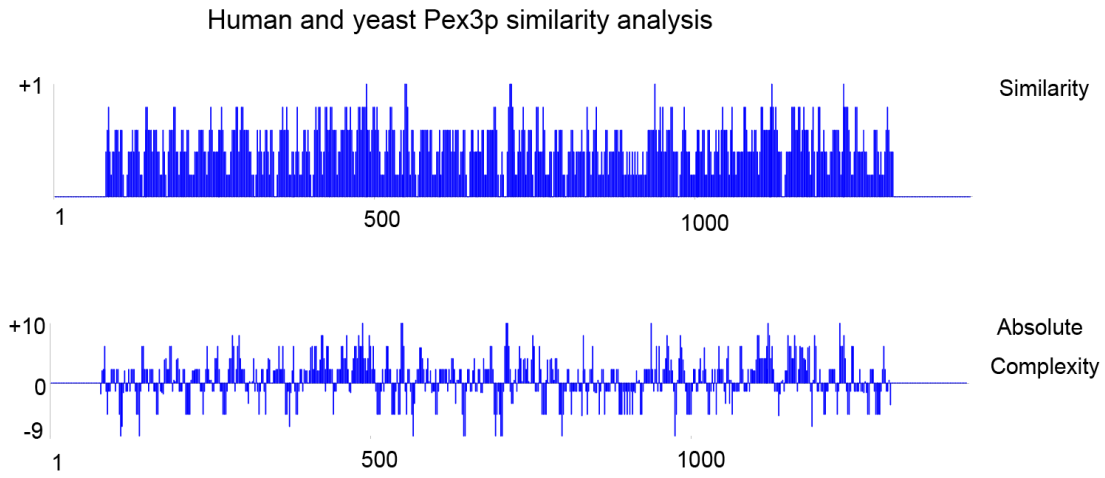


Figure 5.2: Human and yeast Pex3p sequence analysis. The plot shows a 40.3% identity. This figure was generated by aligning and comparing human and yeast *PEX3* nucleotide sequences.

At the time of inception of this project, the structures of human Pex3p and Pex19p had not been resolved. They have since been described, as shown in Chapter 2.12, Page 32. The structures of yeast Pex3p and Pex19p have yet to be resolved using X-ray crystallography or other techniques.

5.2.2 Yeast Strain Selection

Two background strains were selected for this project. The first was: YPH500α (*MATα ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1*) (Sikorski and Hieter 1989). The

second was: BY4741a (*MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1*) (Paciello, de Alteriis et al. 2009). The Sikorski vector compatibility was necessary as the project was originally designed with five constructs in mind. The corresponding Sikorski chromosomal integration vectors were used: *URA3* - pRS306, *TRP1* - pRS304, *LEU2* - pRS305 and *HIS3* - pRS303. *ADE2* is involved in the adenine synthesis pathway in *S. cerevisiae* and when the yeast is grown in absence of the amino acid, it turns a bright shade of red that is easily detectable by eye (Sharma, Kaur et al. 2003).

5.2.3 Project Methodology and Progression

BioBrick parts were assembled into the Sikorski vectors for transformation into yeast. In compliance with BioBrick assembly techniques, all parts had forbidden restriction sites *EcoRI*, *XbaI*, *SpeI* and *PstI* removed by site-directed mutagenesis. All plasmid maps were constructed using Invitrogen Vector NTI 10. This software was also used for all other genetic analysis including sequence trace analysis and restriction site searches. A *PEX3* and *PEX19* double knockout was necessary to avoid having any untagged copies of the membrane protein in the cell.

Sikorski vectors are yeast integration vectors that function via the homologous recombination DNA repair system (Sikorski and Hieter 1989). They were linearized in order to integrate into the genome. The *HIS3*, *TRP1* and *URA3* vectors are BioBrick compatible and can be linearized using the standard 4 enzymes (*EcoRI*, *XbaI*, *SpeI* & *PstI*). However, the LEU2 vector is not, so *BstEII* site was chosen for linearization. The constructs were searched for any other occurrences of this site and none were found, implying no further site directed mutagenesis was necessary. Up to five constructs were planned per strain. Initially the yeast *PEX3* promoter was to be used in the project to direct the transcription of human and yeast *pex3*. Unfortunately, several attempts to clone the *PEX3* promoter constructs using E. coli strains DH5α and TOP10 were unsuccessful. Several attempts were made to clone the construct using *E. coli* strains DH5α and TOP10 without success. This is the only time an *E. coli* cytotoxic part was observed in the duration of the project. A promoter with similar expression levels under Raffinose/oleic acid media was needed so the *PEX8* promoter was chosen.

5.3 BioBrick Part, Construct and Strain Assembly

Many parts were used to construct the DNA assemblies for this project. This section shows progression from individual parts to completed Sikorski vector constructs. The general assembly structure was as follows:

Part → BioBrick assembly → Sikorski vector assembly → Strain

Individual parts were made by designing PCR primers containing the BioBrick-compatible restriction sites and amplifying a region of DNA of interest, as explained in the introduction. Individual parts have a single letter and number designator, such as T1 (terminator 1) or P6 (promoter 6) as shown in Table 5.1. Once parts were assembled, their designation was changed to kZ plus an assembly number. The kZ designation represented parts composed together and selected for cloning into a Sikorski vector. When the Sikorski vector was correctly assembled, it was linearized to allow it to integrate into the genome when transformed. Completed yeast strains were designated KDY followed by a number. Each subsequent transformation incremented the number following the KDY yeast designation.

5.3.1 BioBrick individual parts

Part type	Code	Abbreviation	Name	Length (bp)	BioBrick name
Promoters	P6	pPex3	*S. cerevisiae Pex3* promoter	276	
	P7	pPex8	*S. cerevisiae* Pex8 promoter	241	
	P9	pPex19	*S. cerevisiae* Pex19 promoter	184	
	P11	pHis3	*S. cerevisiae* His3 promoter	225	
Terminators	T1	Adh1t	S. cerevisiae ADH1 terminator	225	L0101
Kozaks	K1	hsKozak	Ira's Human Kozak	18	K0001
	K2	scKozak	S. cerevisiae optimized Kozak	12	
Fluorophores	F1	1xYFP.v	1x Venus YFP	744	E0130
	F2	1xRFP.m	1x mCherry RFP	735	E0110
	F3	1xCFP.c	1x Cerulean CFP	744	E0120
	F4	2xYFP.v	2x Venus YFP	1494	Z0035hc
	F5	2xRFP.m	2x mCherry RFP	1476	Z0037hc
	F6	2xCFP.c	2x Cerulean CFP	1494	Z0541
Genes	G1	*scPEX3*	*PEX3* (*S. cerevisiae*)	1335	
	G2	*scPEX19*	*PEX19* (*S. cerevisiae*)	1062	
	G3	*hsPEX3*	*PEX3* (*Homo sapiens*)	1122	
	G4	*hsPEX19*	*PEX19* (*Homo sapiens*)	900	
Tags	S11	cPTS1	c-Terminal Peroxisomal Targeting Sequence	12	
	S12	scKozak. nPTS2	n- Terminal Peroxisomal Targeting Sequence	48	
Vectors	V1	V0120	V0120 BioBrick Vector	4967/3240	

Table 5.1: Individual BioBrick parts used in the project. Parts with BioBrick names are in the registry of standard biological parts, while parts with no BioBrick names were used in the project and not submitted.

5.3.2 BioBrick kZ part to assembly summary

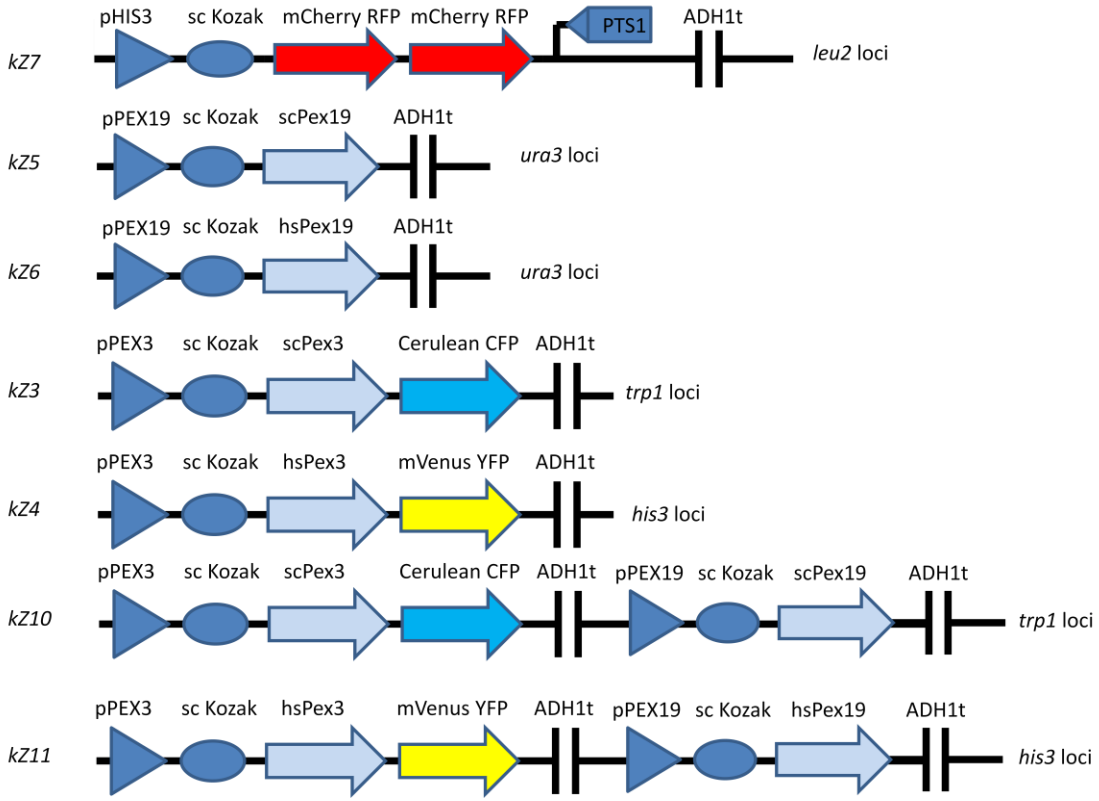


Figure 5.3: Schematic of constructs assembled during the cloning phase of the project.

Figure 5.3 shows a schematic of the BioBrick constructs assembled from individual parts shown in Table 5.1. These constructs were designed to be used in the YPH500α and BY4741a background strains. Two strategies needed to be developed for this part of the assembly. The first involved using three constructs, kZ7, kZ10 and kZ11. This strategy was based around having two constructs on the kZ10 and kZ11 assemblies: Pex3p-fluorophore and Pex19p. This would have limited the number of serial transformations to three for the main experimental strain; the RFP-PTS assembly, the CFP and YFP constructs. However, it was not known if having a single terminator would produce plasmid instability (during cloning) or read-through problems, so a secondary strategy was also devised. This strategy called for using kZ3, kZ4, kZ5, kZ7 and a version of kZ6 that was planned to be constructed using genetecin selection. This strategy would have used all four Sikorski vectors and a geneticin selection for the final

strain. A cloning strategy involving 5 sequential transformations was not ideal so the first strategy was employed.

5.3.3 Strain construction in the YPH500α and BY4741a backgrounds

Strain name	Genotype (YPH500α background)
YPH500α (wt)	*MAT α ura3-52 lys2-801_amber ade2-101_ochre trp1- Δ63 his3- Δ200 leu2- Δ1*
KDY1	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 leu2- Δ1 ura3-52 pex3-Δ*
KDY2	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 leu2- Δ1 ura3-52 pex19-Δ*
KDY3	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 leu2- Δ1 ura3-52 pex3-Δ pex19-Δ*
KDY4	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 ura3-52,* LEU2::pHIS3-RFP-RFP-PTS1
KDY5	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 ura3-52,* LEU2::pPEX8-RFP-RFP-PTS1
KDY6	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 ura3-52,* LEU2::pPEX19-RFP-RFP-PTS1
KDY7	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 ura3-52 pex3-Δ ,* LEU2::pHIS3-RFP-RFP-PTS1
KDY10	YPH500 *MAT α ade2-101_ochre trp1- Δ63 his3- Δ200 ura3-52 pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1
KDY19	YPH500 *MAT α ade2-101_ochre leu2- Δ1 pex3-Δ pex19-Δ* TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19

Table 5.2: KDY strain construction details. Strains KDY1 – KDY3 were constructed to test the PEX3Δ, PEX19 Δ and PEX3 Δ, PEX19 Δ knockouts. *The main experimental strain (KDY19) was constructed without the LEU2::pHIS3-RFP-PTS1 construct.

Table 5.2 shows the progression of the yeast strains. Strains KDY1 – KDY3 are the knockout strains. The KDY4 – KDY6 strains were built to find the best promoter to use with the RFP-PTS1 construct between *HIS3*, *PEX8* and *PEX19*. KDY7 was planned as a control to test that *PEX3* had been successfully knocked out. KDY8 tests the successful rescue of peroxisome import function using human yeast Pex3p can rescue the knockout. It was also to be used to test the performance of CFP as a fluorophore and whether the *PEX3* promoter is a strong enough promoter to see any activity. KDY9 has a similar function with the yeast *PEX3* rescue and the YFP fluorophore. KDY10 was designed to verify the deletion of *PEX19*.

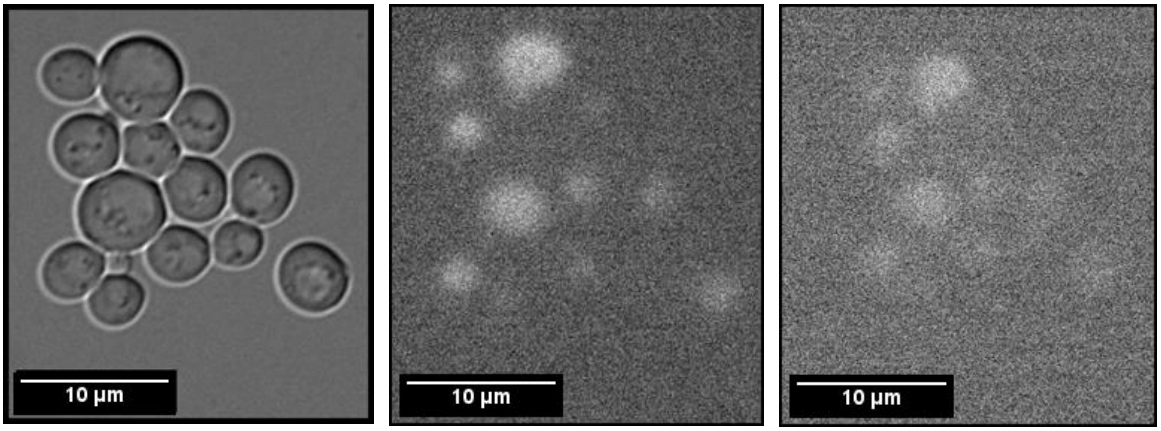
Strain Name	Genotype (BY4741a background)
KJY1	wt BY4741 *MATa* LEU2::pHIS3-RFP-RFP-PTS1
KJY2	BY4741 *MATa , pex3-Δ,* URA2::pHIS3-RFP-RFP-PTS1
KJY3	BY4741 *MATa , pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1
KJY4	BY4741 *MATa , pex3-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP
KJY7	BY4741 *MATa , pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1 TRP1::pPEX19-scPEX19
KJY8	BY4741 *MATa , pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t -pPEX19-scPEX19
KJY9	BY4741 *MATa , pex3-Δ, pex19-Δ* LEU2::pHIS3-RFP-RFP-PTS1, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19
KJY11	BY4741 *MATa , pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19
KJY14	BY4741 *MATa , pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-hsPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19
KJY15	BY4741 *MATa , pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-scPEX3-YFP-ADH1t-pPEX19-scPEX19
KJY10	BY4741 *MATa , pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19

Table 5.3: KJY strain construction details. Strains KDY1 – KDY3 were constructed to test the PEX3Δ, PEX19 Δ and PEX3 Δ, PEX19 Δ knockouts.

Table 5.3 shows the genotypes for the KJY strain collection. As the BY4741 background was capable of importing proteins into the peroxisome using the PTS1 import tag, it was used for the majority of experiments discussed later in the chapter. The KJY10 strain is the main experimental strain. It used the *PEX3* and *PEX19* double knockout background and possesses the RFP-PTS1 import tag, the human Pex3p-CFP and yeast Pex3p-YFP constructs.

5.3.4 Media Formulation

Peroxisomes are known to be repressed (although cells maintain a minimum number of the organelles) when grown on glucose and induced when grown on fatty acids and raffinose as a carbon source. To this end, several media were researched and developed using glucose, raffinose, potassium acetate, glycerol and oleic acid as carbon sources. SA and YNO did not meet with much success due to the high background fluorescence and consequent low signal to noise ratio. YEDP/YPDA and SD often showed high background fluorescence, producing low quality images that made automated image analysis difficult. YPGO and its variants, with the exception of STYO/STYO-2, produced better results.



A) **B)** **C)**

Figure 5.4: (YPH500α *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) yeast cells imaged using differential interference contrast microscopy (DIC), B) cells imaged using the YFP channel (mVenus) C) yeast cells imaged using the CFP channel (Cerulean). This figure shows an example of insufficient fluorescence signal. The strain was grown in SD media for 24h then imaged while live. The strain contains human and yeast Pex3-YFP and yeast-CFP peroxisomes.

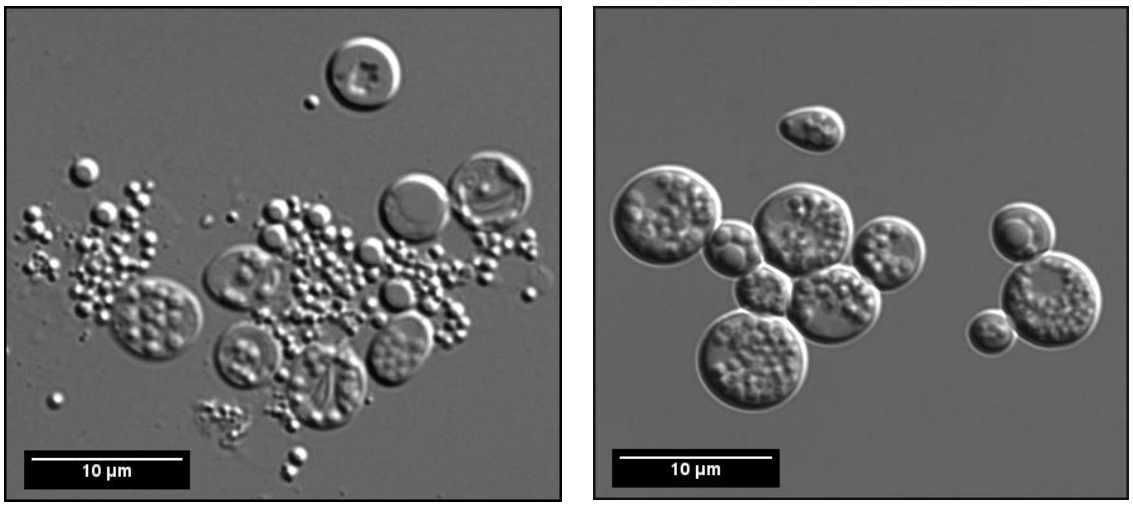
Figure 5.4, shows yeast grown and imaged live in SD media. Figure 5.4 A shows a DIC image of the strain while B and C show the YFP and CFP fluorescence. The graining effect seen in these images indicates a very weak fluorescence signal. Greater induction of the peroxisomes was necessary and would be achieved though changes in the composition of the growth media. However, B and C show vacuoles that appear fluorescent from an accumulation of a purine intermediary in the adenine biosynthesis pathway in the *ADE2* deletion strain, not peroxisomes.

A number of yeast growth media were formulated and tested during the media exploration phase of the project. Often, the results were either a minor incremental change in the quality of the images, or no change at all. All media recipes have been given, but only successful tests will be shown in the results section.

5.3.5 Mounting Cells for Microscopy

Initially, strains grown in SD/YEPD/YPDA were mounted in double distilled water (ddH_20) on glass slides for microscopy. When strains grown in YPGO or SCGO media were mounted using

ddH_20, cell lysis was often observed as illustrated in Figure 5.5 A. Oleic acid media contains a surfactant such as Tergitol NP-40, Tween40 or Triton x100 to disperse the oleic acid.



A) **B)**

Figure 5.5: (YPH500α wt: *MATα ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1)* A) Lysed cells subject to Osmotic shock due to incorrect mounting media. B) Cells grown in SCGO and imaged live with correct osmotic pressure osmolarity during live imaging.

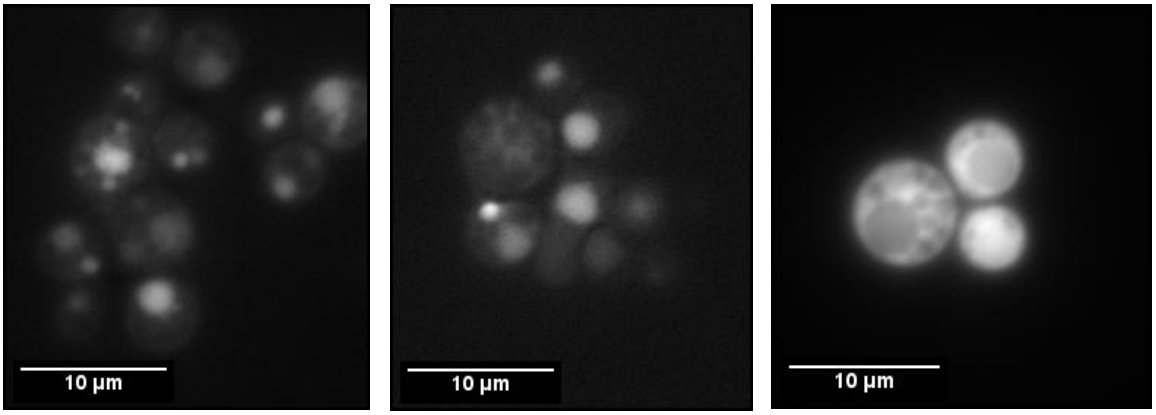
To prevent lysis, a mounting media was formulated to mimic the osmolarity of the surfactants used in the oleate growth media. The mounting media consisted of the same concentrations of oleic acid and surfactant as found in the growth media, without any amino acids or yeast nitrogen base. Cells mounted using this new formulated media can be seen in B Figure 5.5. Once created, this media was used in all subsequent live cell imaging for cells grown in oleic acid media. Cells that were paraformaldehyde fixed were imaged in a solution of ddH_20 containing 1.2 M sorbitol and 0.2 M potassium phosphate.

There were several problems with imaging live cells. The peroxisomes do not have a strong fluorescence signal when tagged with CFP and require long acquisition times (up to 2 seconds) when an electron-multiplying charge coupled device (EMCCD) camera is not employed. Imaging a single focal plane for two seconds can result in unclear images as the peroxisomes can move in and out-of-focus. As the peroxisomes are not fixed within the cell, acquiring multiple images in a z-stack does not solve the problem. A solution was to crosslink the cells using a

paraformaldehyde fixation protocol prior to imaging as described in Chapter 3. The cells could now be fixed and imaged at a later date, instead of a two hour window and the peroxisomes no longer travel across the focal plane. The cells were then mounted in the KPO_4/sorbitol mounting media. This process yielded much brighter and clearer images at the expense of killing the cells.

5.3.6 Initial Fluorescence Imaging Experiments

The strain construction was separated into two major sections. The *PEX3-PEX19* double knockouts were constructed in parallel to the RFP-PTS1 test strains. Once the RFP-PTS1 construct was transformed into the *PEX3-PEX19* double knockout, the other constructs could be inserted and there would always be a positive control that the peroxisomes were functional. However, the imaging results influenced the design and first the correct promoter was selected to drive the RFP fluorophore at an appropriate expression level.



A) **B)** **C)**

Figure 5.6: A) *HIS3* promoter RFP (mCherry)(YPH500 *MATα ade2-101_ochre trp1-Δ63 his3-Δ200 ura3-52,* LEU2::pHIS3-RFP-RFP-PTS1) B) *PEX8* promoter RFP (mCherry)(YPH500 *MATα ade2-101_ochre trp1-Δ63 his3-Δ200 ura3-52,* LEU2::pPEX8-RFP-RFP-PTS1) C) *PEX19* promoter RFP (mCherry)(YPH500 *MATα ade2-101_ochre trp1-Δ63 his3-Δ200 ura3-52,* LEU2::pPEX19-RFP-RFP-PTS1). The strains were grown in YEPD media for 24h and the microscopy was performed while the cells were live.

Figure 5.6 shows micrographs of the strains hosting the *HIS3*, *PEX8* and *PEX19* promoter constructs expressing RFP. Figure 5.6 A (*HIS3* promoter) shows some highly fluorescent circles thought to be vacuolar localization of RFP with some smaller regions that appear to show peroxisomal localization. B (*PEX8* promoter) shows only vacuolar localization and (*PEX19*

promoter) shows only cytoplasmic localization. Based on these preliminary images, the decision was made to use the *HIS3* promoter to express RFP-PTS1. The images in Figure 5.6 appear to show a different localization of the RFP based on the promoter. The observation was made, but not investigated as this stage in the project was purely about strain construction.

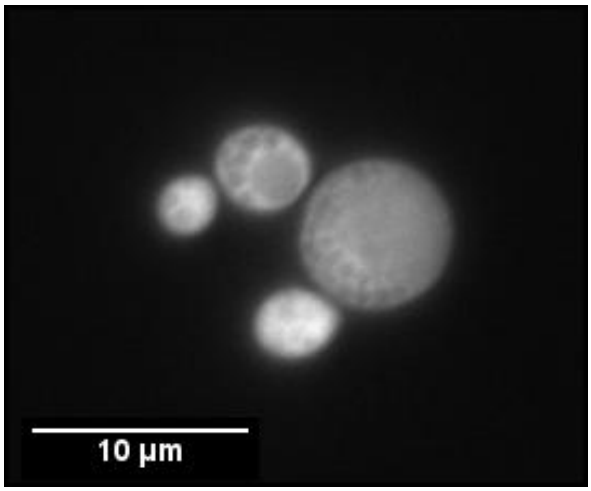


Figure 5.7: (YPH500 *MATα ade2-101_ochre trp1-Δ63 his3-Δ200 ura3-52 ,* LEU2::pHIS3-RFP-RFP-PTS1) RFP-PTS1 with cytoplasmic RFP (mCherry). The cells were grown in YPDA medium for 24 hours and the cells were imaged live.

Once *HIS3* was selected, the construct was imaged again using the YPGO media to fully induce the peroxisomes. Figure 5.7 is from this experiment and shows cytoplasmically located RFP. The RFP was characterized as cytoplasmic as there the vacuoles do not appear to have fluorescence, nor do any of the other subcellular organelles. The RFP appears brighter in the daughter cells, but the resolution does not allow for identification of subcellular compartments. This experiment led to a shift in focus to construction of the double knockout strain and the Pex3p-fluorophore expression cassettes.

5.3.7 PEX3-PEX19 Double Knockout Strain Construction

In parallel to the RFP-PTS1 promoter selection and the BioBrick part assembly, the double knockouts were constructed. This process involved designing oligonucleotide primers containing a 50-75 base pair region of DNA for homologous recombination of a selection cassette in the loci of the gene to be removed. Constructs were assembled by PCR of the

selection cassette with the homologous recombination primers. Both *PEX3Δ* and *PEX19Δ* single knockout strains were made at the same time. The knockouts were made using the LoxP flippase system so that the phleomycin-resistance selectable marker could be subsequently removed and re-used; the protocol is given in Chapter 3.4. Both knockouts were made without using two selectable markers, as it was necessary to keep up to five markers available. This was necessary incase assembling Pex3 and Pex19 into the same vector failed. The PUG72 (URA3) & PUG73 (LEU2) EUROSCARF vectors were used initially, but several attempts failed. The PUG66 phleomycin resistance vector was eventually chosen as it had previously been used with success in the lab, while the other variants had not (C. Kennedy, personal communication).

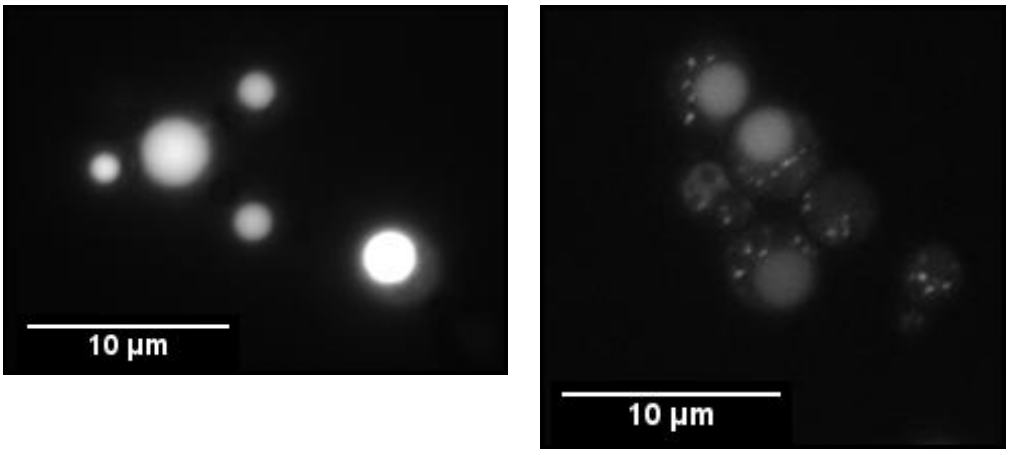
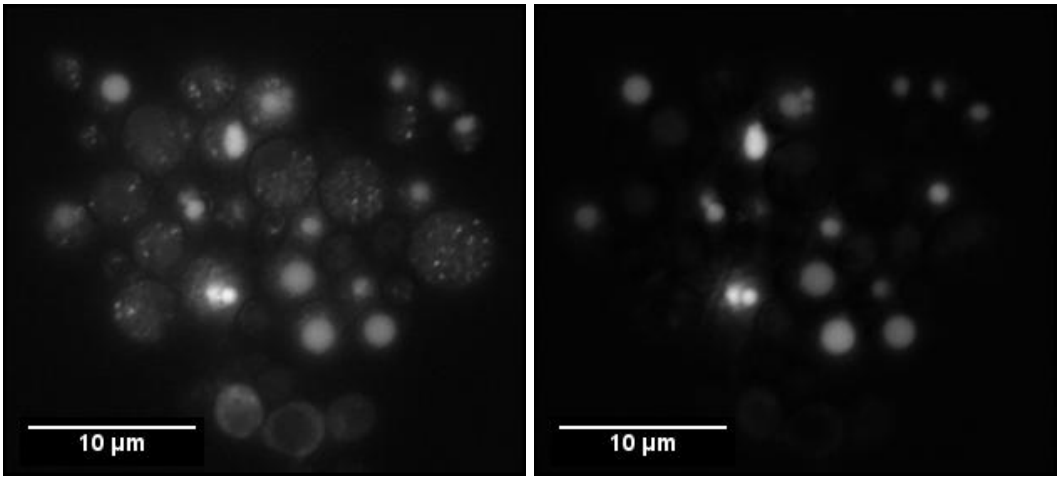


Figure 5.8: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* trp1::ppex3-scpex3-cfp-adh1t-ppex19-scpex19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) CFP imaging - KDY13 grown in YFGP (Cerulean) B) KDY12 YFP peroxisomes and vacuolar localization (mVenus). The cells were grown in YPDA medium for 24 hours, peroxisomes were induced for 24 hours in YPGO medium and the cells were imaged live.

Figure 5.8 shows the first images taken with the CFP and YFP yeast Pex3p tagged constructs. Figure 5.8 A shows extremely bright, single circles in each cell. Figure 5.8 B has the same circles at a lower intensity but also has smaller dots in each cell. B shows the first observation of peroxisomes. It was assumed that the same observations couldn't be made in A as the circles were much brighter than the peroxisomes consequently obscuring them from view. These circles were vacuolar localization from the *ADE2* deletion (Weisman, Bacallao et al. 1987).

Mutations in *ADE2* lead to the accumulation of purine precursors in the vacuole, which produces the red pigmentation and fluorescence (Zonneveld and van der Zanden 1995). Fluorescence was previously thought to only exist on the RFP channel.



A) **B)**

Figure 5.9: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* trp1::ppex3-scpex3-cfp-adh1t-ppex19-scpex19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) yeast expressing yeast Pex3p-YFP with peroxisomal and cytoplasmic localization (mVenus) B) human Pex3p-CFP with cytoplasmic localization (Cerulean). The cells were grown in YPDA medium for 24 hours, peroxisomes were induced for 24 hours in YPGO medium and the cells were imaged live.

Figure 5.9 shows a yeast strain grown in YEPD for 24 h then transferred into YPGO for a further 24 h and imaged while live. Figure 5.9 A shows YFP-human peroxisomes and vacuolar localization while Figure 5.9 B shows only vacuolar localization. The vacuolar problem continued to occur, so a solution needed to be found. However, a new observation could now be made. The vacuolar fluorescence was approximately the same intensity in both images, so it was apparent that B did not contain any peroxisomes in the cells. Two problems now needed solutions, the vacuolar fluorescence and the absence of yeast CFP peroxisomes.

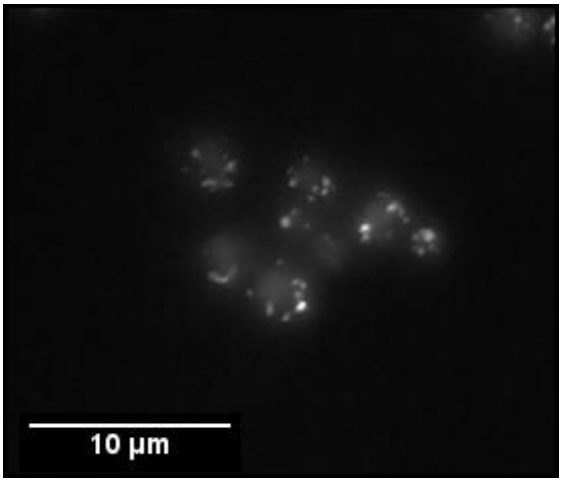
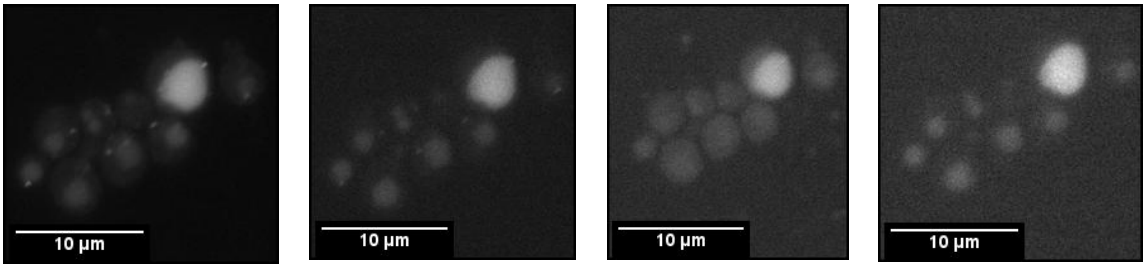


Figure 5.10: Pex3p-GFP from the GFP collection using Aequorea victoria GFP (Tsien 1998). The cells were grown in YPDA medium for 24 hours, peroxisomes were induced for 24 hours in YPGO medium and the cells were imaged live.

Figure 5.10 shows a fluorescence control with known GFP tagged peroxisomes. This figure shows Pex3p-GFP strain from the GFP collection created by Erin O'Shea (Huh, Falvo et al. 2003). When compared to Figure 5.10, the images in Figure 5.6 appear to all show vacuolar localization to some degree. The Pex3p-GFP construct was used in the GFP collection work as a co-stain to identify other peroxins. In this work, it was used it as a control to show an example of clear, discrete, membrane tagged peroxisomes. In Figure 5.10, an observation can be made that there is no vacuolar localization and the peroxisomes can be clearly seen as discrete white objects. The background strain phenotype includes an *ADE2* deletion. This deletion causes an interruption of the adenine synthesis pathway producing a pigment that is red in color. It was postulated that this otherwise useful phenotype was interfering with the microscopy.



A) **B)** **C)** **D)**

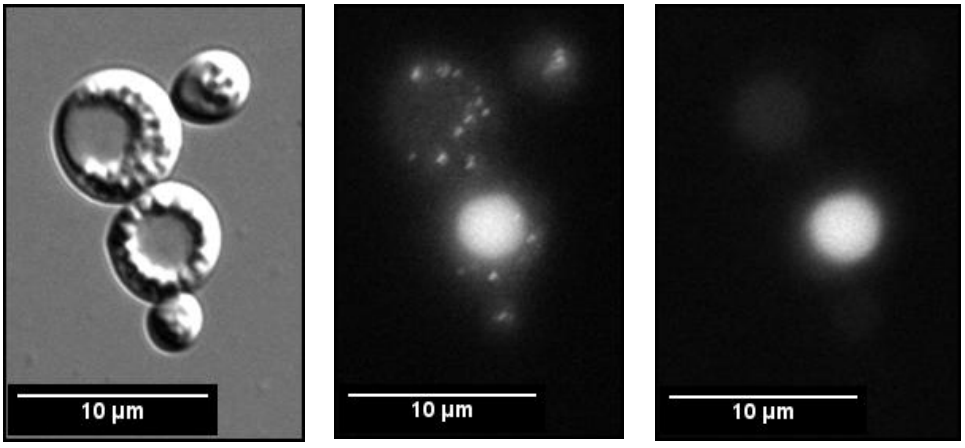
Figure 5.11: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* trp1::ppex3-scpex3-cfp-adh1t-ppex19-scpex19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) FITC (mVenus) B) YFP (mVenus) C) CFP (no fluorophore) D) RFP (no fluorophore). Images show the KDY3-KX9-KX10 strain grown overnight in YEPD, induced for 7 hours in SD media and imaged live.

Figure 5.11 shows a strain containing human-YFP and yeast-CFP peroxisomes. In an attempt to visualize any interference, all four fluorescence channels were imaged after the yeast had been grown in synthetic dextrose (SD) media. The SD media lacks certain fluorescent elements found in yeast extract, eliminating fluorescence from the media itself. All images show vacuolar localization, but Figure 5.11 A is the only image showing peroxisomes. This highlighted interference from the media lacking adenine and also a lack of CFP-fluorophores despite the presence of the KX10 (yeast Pex3p-CFP) construct. These images illustrate the choice to select FITC instead of the YFP channel. The YFP peroxisomes can clearly be seen in Figure 5.11 A (FITC), but not in B (YFP). An unfortunate and common phenotype is the pigmentation with fluorescence in most of the visible spectrum, causing the vacuole to appear as an intensely fluorescing object under microscopy (Sharma, Kaur et al. 2003).

Vacuolar fluorescence was noted during most of the previous experiments but the cause was assumed to be interference from the YEPD medium. Yeast extract used in the medium is known to have fluorescent elements that can interfere with microscopy. To reduce this effect, synthetic medium was used where possible. When the cause of this effect was attributed to the *ADE2* deletion in the strain itself, adenine was added to the media in a successful attempt to reduce background interference.

5.3.8 CFP – YFP Strain Re-construction

When both Pex3p tagging constructs had been transformed into the strain, they were verified by colony PCR and subsequently by imaging.

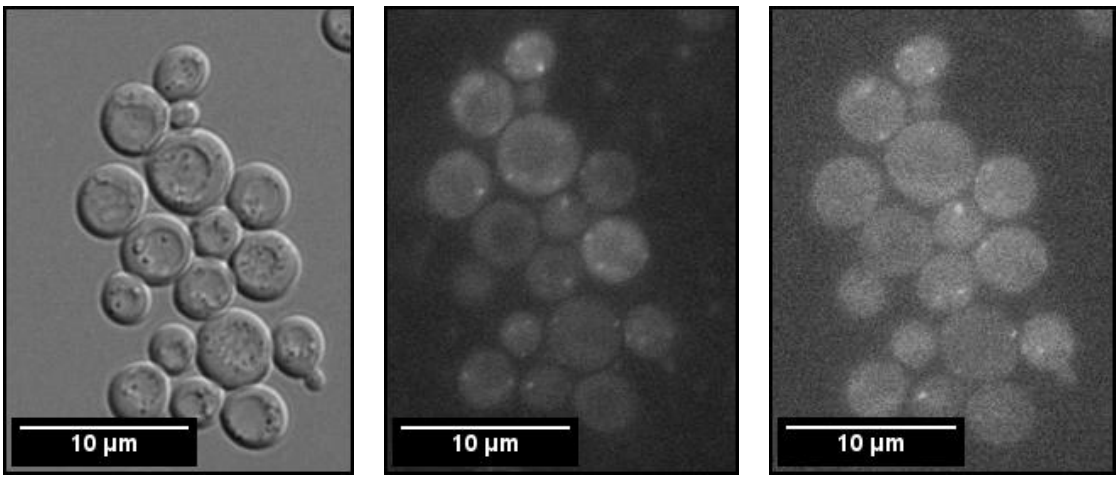


A) **B)** **C)**

Figure 5.12: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* trp1::ppex3-scpex3-cfp-adh1t-ppex19-scpex19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) DIC, B) YFP-human peroxisomes (mVenus) C) CFP channel without peroxisomes (no fluorophore). The strain was grown in SD media for 24h then transferred to YGPO for peroxisome induction for 24 hours. The cells were imaged live.

Figure 5.12 shows the strain expressing only YFP peroxisomes. When several digests and analytical gels were run on the assemblies in question, one of the constructs was 200 bp smaller than expected. Subsequent analysis showed the kZ3 construct was missing the promoter driving expression of Pex3p-CFP, hereby explaining the lack of CFP yeast peroxisomes.

Once the discrepancy had been discovered, the construct was rebuilt and transformed into the working YFP strain, creating the new KDY3-KX9-KX10-KX9_URA3 strain. This long designation is explained by the strain having two copies of the KX9 construct. The original malfunctioning KX9 was supplemented with a new working copy using a different Sikorski vector. The strain contained functional copies of CFP-hs*PEX3* and YFP-sc*PEX3*, but possessed two functional copies of the yeast *PEX19*.



A) **B)** **C)**

Figure 5.13: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) DIC, B) CFP-human peroxisomes (Cerulean) C) YFP-yeast peroxisomes (mVenus). These images represent the first successful CFP peroxisome images. Strain: KDY3-KX9-KX10-KX9_URA3. The cells were grown in SD medium overnight to stationary phase, induced in YPGO for 24 h and imaged live.

Figure 5.13 shows the human and yeast peroxisomes in the same strain. The cells were imaged live and show a low but visible peroxisome signal on both the CFP and YFP channels. Adenine was used in the media so the cells show no trace of vacuolar fluorescence.

5.4 Z-stacks and Automated Image Processing using 3d deconvolution

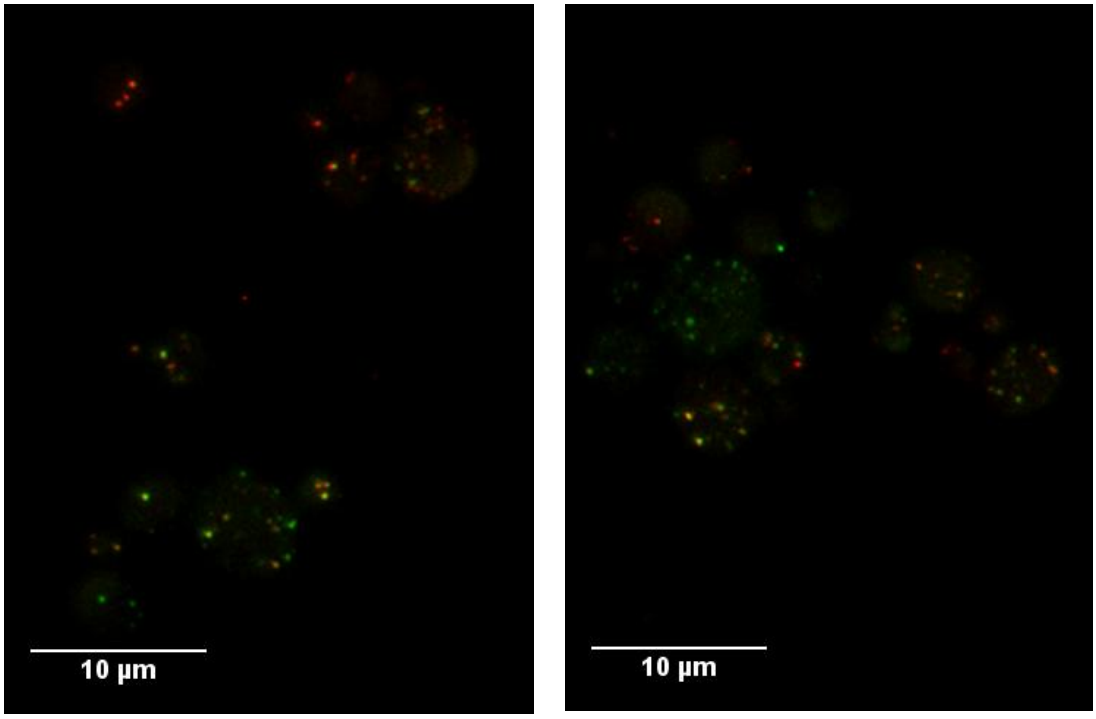
DeltaVision uses its own proprietary acquisition software, but a choice of suites was available for image analysis. Initially, MetaMorph was used due to familiarity. However, poor scripting support led to the eventual use of Image Pro for post processing and AutoQuant X for 3d deconvolutions.

5.4.1 Z-stack details

Once the exposure times and transmitted light percentages were set, the Z-stack depth was set. For consistency, this value is was always 10.20 µm. A Z-step of 0.20 µm was as close as possible to the XY camera pixel size of 0.12760 µm. The voxels (3 dimensional pixels) are effectively rectangular in shape when looked at from the 3d perspective, but 0.20 µm is the minimum accurate resolution of the microscope stage. Once all the acquisition parameters were

verified and re-verified and the experimental settings were saved, the microscope automatically imaged the ~50 Z-slices over four wavelengths. This process was repeated up to ten more times without changing any settings other than the XY position on the microscope slide. Once a sufficient number of images were acquired, the process was repeated a last time on a region of the slide with no cells to produce the blank. This image acquisition protocol was repeated with all samples and controls. Experiments were conducted using fewer z-slices, but problems were encountered when there weren't enough images above and below the regions of interest to successfully apply the point spread function (PSF) during deconvolution. This resulted in the entire stack of images being lost for processing purposes. This technique was developed as once the first z-stack was acquired, the imaging protocol was not changed from region to region on the slide as a greater number of z-slices would account for any potential variations in thickness.

5.5 3d Deconvolved Images using the YPH500α Strain



A) **B)**

Figure 5.14: (YPH500 *MATα ade2-101_ochre leu2-Δ1 ura3-52 pex3-Δ pex19-Δ* TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) False color CFP-YFP (cerulean & mVenus) overlay, B) False color CFP-YFP (cerulean & mVenus) overlay 2. The strain shown is KDY3-KX9-KX10-KX9_URA3. These false color images were acquired using a strain that was grown to saturation in YPDA, induced in YPGO for 24 h and paraformaldehyde fixed. CFP is colored green and the YFP, red to improve contrast.

Figure 5.14 shows false color images of the working CFP (green) and YFP (red) strain. To produce the image, CFP and YFP channels were imaged in a Z-stack and a 3d deconvolution was performed. These deconvolved Z-stacks were compressed into one image and then both channels were added together. Figure 5.14 A shows red peroxisomes in the top left hand corner of the image and green ones in the bottom left. Figure 5.14 B shows more red and green peroxisomes with an added cell containing yellow peroxisomes. This could indicate that the cells have entirely CFP, YFP or mixed peroxisomes. This is a rough observation as a bright-field image was not taken in this image-set so cell delineation is difficult. However, both images show some degree of bimodality in the peroxisome populations.

5.5.1 RFP-cell Outline Images

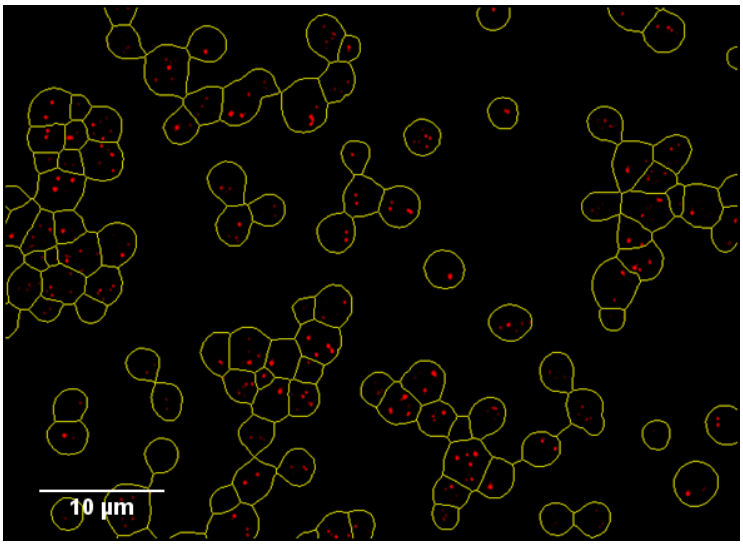


Figure 5.15: (BY4741 *MAT*a LEU2::pHIS3-RFP-RFP-PTS1) RFP peroxisomes (mCherry) with cell outlines after image processing. Cells were grown in YPDA overnight, induced in YPGO for 24 hours and imaged fixed.

The bright-field out-of-focus imaging was developed to produce a protocol to delineate cell outlines. Figure 5.15 shows cell separation based on the out-of-focus bright-field image with a strain containing RFP-PTS1 peroxisomes. This cell delineation analysis technique was incorporated into the first module in the Image Pro script.

5.5.2 CFP-YFP cell outline

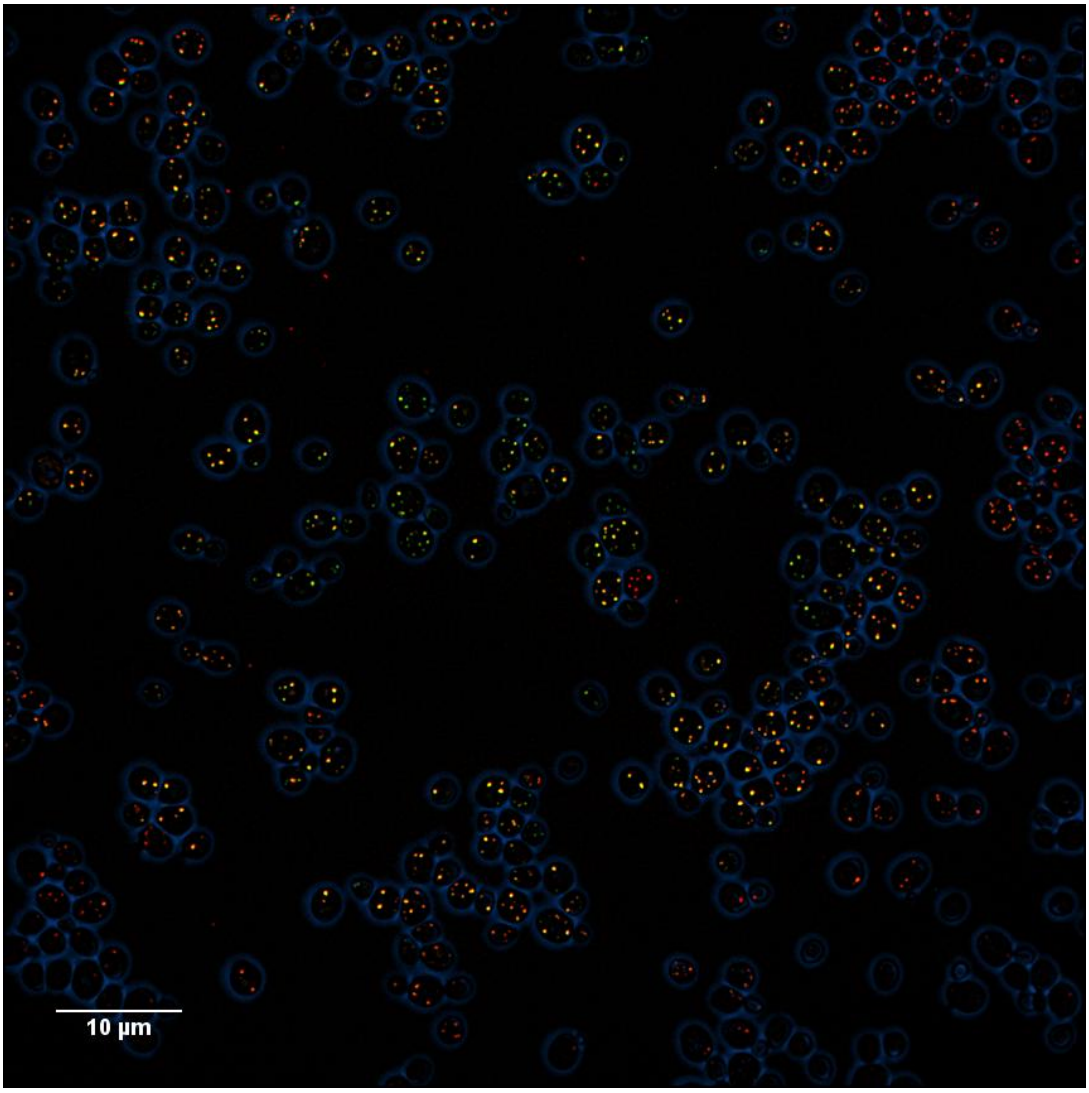


Figure 5.16: (YPH500 *MAT*α *ade2-101_ochre leu2-Δ1 pex3-Δ pex19-Δ* TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) False color composite image of CFP (Cerulean) and YFP (mVenus) peroxisomes with blue cell outlines. Cells were grown in YPDA overnight, induced in YPGO for 24 hours and imaged fixed.

The best image produced in the microscopy evolution using the KDY19 (YPH500 *MAT*α ade2-101_ochre leu2-Δ1 pex3-Δ pex19-Δ TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) strain is shown in Figure 5.16. This image is made from three channels: CFP, YFP and out-of-focus bright-field. As in earlier images, the CFP channel is colored green and the YFP channel, red to enhance contrast. The

image shows entirely green peroxisomes and entirely red peroxisomes, indicating some degree of bimodality. However, upon closer examination the green peroxisomes appear predominantly in the centre of the image while the red ones appear towards the outer edges. At the very extreme edges of the image, no peroxisomes are present in the cells. To further investigate this effect, an intensity plot was made from one diametrically opposite corner to the other where each point was assigned a numerical value. The average of this value appeared to increase in the centre and decrease towards the edges on both the CFP and YFP plots. The microscope is set up so that the illumination in the field of view is as consistent as possible to avoid introducing this kind of bias. Various components of the microscope were examined and the xenon illumination bulb was determined to be faulty. The silver coating had partially burning off, likely due to insufficient cooling, and produced an uneven illumination and a halo effect on the image. This effect explained the bimodality and unfortunately invalidated the quantitative analysis due to the illumination bias.

5.6 Construction and Development of the Control and Main Experimental Strains

This section describes work with the BY4741a control strains, which were analyzed because they contained a functional RFP-PTS1 construct that was used as a positive control for peroxisome identification during the algorithmic processing. Each control strain has genotype information, images of that strain and explanation of the images. The construction of the project was structured to show the endogenous peroxisomes functioning correctly in the wild-type background strain. From this starting point, the experimental strains are shown not to have functioning peroxisomes when *PEX3* and *PEX19* were individually knocked out. These genes were removed so that no copies of *PEX19* or untagged copies of *PEX3* exist in the genome. Subsequently, function of *PEX3* and *PEX19* was rescued with the human or yeast tagged equivalent peroxin.

The rescues were demonstrated in the double *PEX3* and *PEX19* knockout, first with the yeast construct and then with the human construct. The experiments were designed so that both *PEX3* and *PEX19* genes are on the same plasmid, to allow for only one transformation per species rescue. In the current constructs, the human Pex3p membrane protein was tagged with a cyan fluorescent protein and the yeast Pex3p, a yellow one. Once these species specific rescues were

demonstrated, they are combined in the same strain. In order to prove that this expected bimodal effect is more than just an artifact of the imaging, numerical results are demonstrated showing CFP/YFP scatter plots and corresponding bimodal histogram in the next Chapter.

5.6.1 Control Strain Information Layout

The control strains are separated into two categories: the rescue controls and the CFP/YFP controls. The rescue controls show successful construction of the single and double knockouts while the CFP/YFP controls will show the interactions of all three fluorophores. Note that human has been abbreviated to hs and yeast to sc.

5.6.2 Control Strain Example Microscopy

Several different images of the same cells are shown in the results section. Figure 5.17 shows an example montage of a strain with three fluorophores.

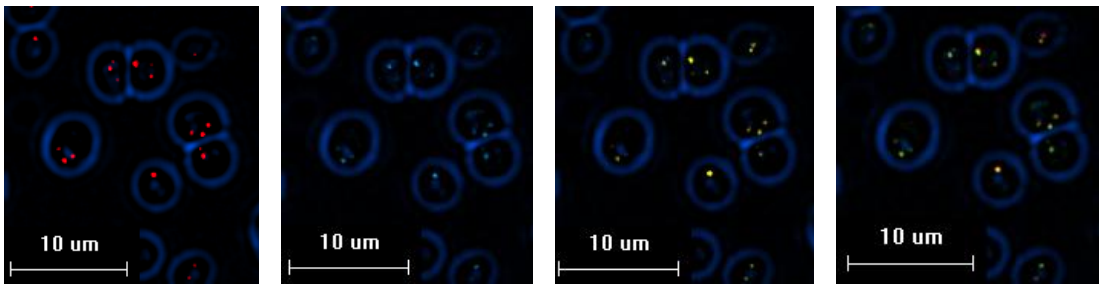


Figure 5.17: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-scPEX3-YFP-ADH1t-pPEX19-scPEX19) A) RFP (mCherry) B) CFP (Cerulean) C) YFP (mVenus) D) CFP-YFP overlay (Cerulean and mVenus). These images show the example color system used with composite images.

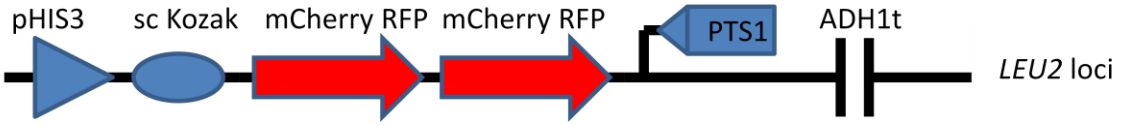
When showing the cell outlines and a single fluorophore, peroxisomes are shown in their fluorescent color (red, yellow or cyan). When showing the cell outlines and two fluorophores, a false color red green image is displayed. In this case, CFP (470 nm) will be shown in green and YFP (535 nm) in red. Manipulating the colors was necessary to increase the contrast in the image to better show the difference between the two channels.

5.7 Rescue Controls

The basic outline for this section of the project was to show the strain information and images of the controls. All cells in this section were grown in YEPD for 24 h, then grown in synthetic raffinose media for 24 h to induce peroxisomes. All strains were paraformaldehyde fixed and stored at 4 ^{o}C until imaged, within a maximum period of 48 hours. Each image is representative of the 4 – 7 that were acquired of the same cells on the slide. As the image processing for quantitative data and the image processing for display requires slightly different protocols, not all images were processed for visualization, but all images were processed for data.

5.7.1 Strain 1: KJY1 – RFP-PTS1 Peroxisome Function Control

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1*



This strain is a control for both the RFP-PTS1 import construct and the endogenous peroxisomes. The presence of peroxisomes indicates the strain is functionally capable of producing organelles without any modifications. Observing peroxisomes indicates the RFP-PTS1 construct is functional.

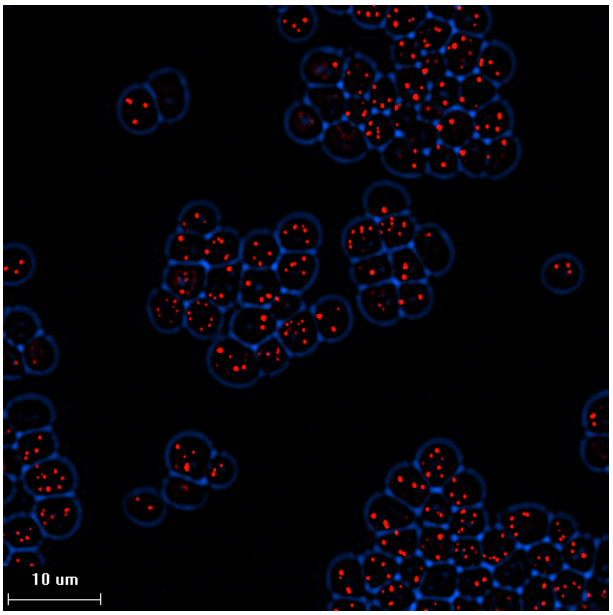


Figure 5.18: (wt BY4741 *MAT*a LEU2::pHIS3-RFP-RFP-PTS1) with RFP-PTS1 tagged peroxisomes (mCherry). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.18 shows how the peroxisomes look with the original *PEX3* and *PEX19* genes, before knocking out the endogenous copies of these genes. The image shows presence of peroxisomes, indicating the strain already has the organelles, the construct is expressed and the RFP-PTS1 tag is functioning correctly. This background strain is therefore able to express peroxisomes and import tagged proteins to demonstrate functionality.

5.7.2 Strain 2: KJY2 – Pex3p Knockout Control

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ*

KJY2 is the first of the knockout strains and is missing *PEX3*. It has the RFP construct to show peroxisome import. As *PEX3* was knocked out, the RFP peroxisomes shown in the previous strain are no longer be present.

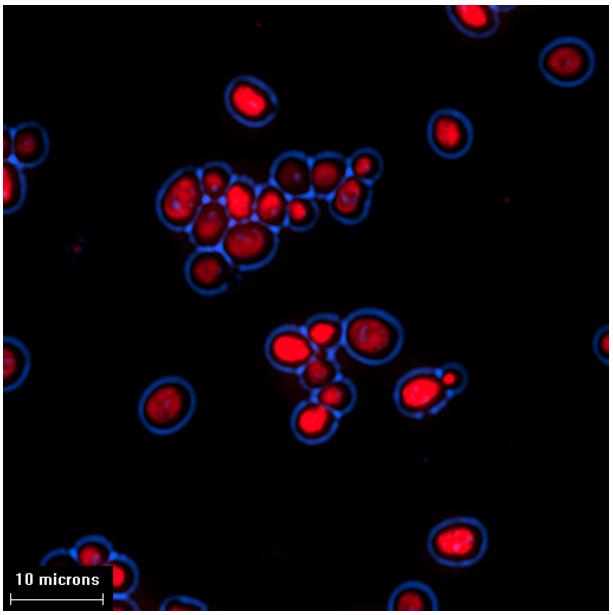


Figure 5.19: (BY4741 *MATa*, *pex3-Δ,* URA2::pHIS3-RFP-RFP-PTS1) Absence of peroxisomes with only cytoplasmic localization of RFP (mCherry). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.19 shows only diffuse cytoplasmically localized RFP. As this strain has the same RFP-PTS1 construct shown to be functional in KJY1, it is clear *PEX3* function has been lost.

5.7.3 Strain 3: KJY3 – Pex19p Knockout Control

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ*

This strain has the other essential peroxin, Pex19p, knocked out. This strain has the RFP-PTS1 construct to show peroxisome import function. As *PEX19* was knocked out, the RFP construct that would usually localize to the peroxisomes now appears diffused in the cytoplasm.

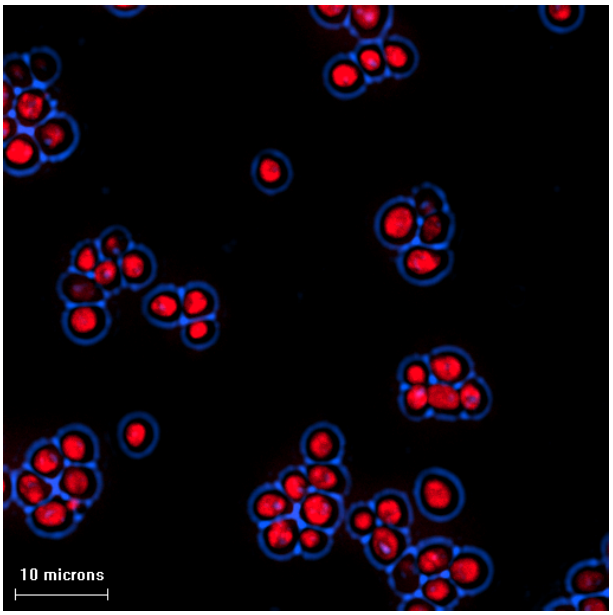
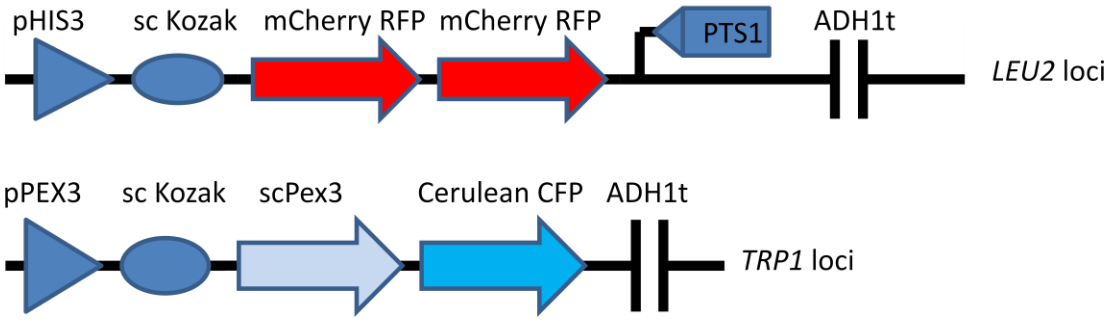


Figure 5.20: (BY4741 *MATa*, *pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1) Absence of peroxisomes and cytoplasmic localization of RFP(mCherry). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

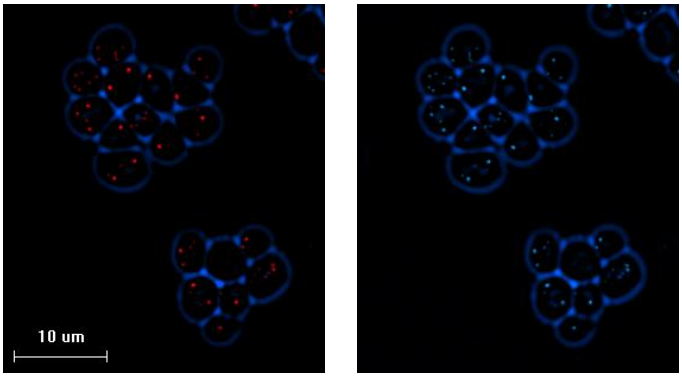
Figure 5.20, shows a lack of distinct peroxisomes, as was also seen in Figure 5.19. From this similarity, we can conclude that the cells are unable to make functional peroxisomes when lacking *PEX19*.

5.7.4 Strain 4: KJY4 – Yeast Pex3p Rescue Control

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ*



The KJY4 strain uses the single *PEX3* knockout to test the yeast *PEX3* rescue function. Two constructs were transformed into this strain. The first is the RFP-PTS1 to verify peroxisome import function. The second is the yeast *PEX3*-CFP construct. As this was a *PEX3* rescue experiment, the original *PEX19* has not been removed.

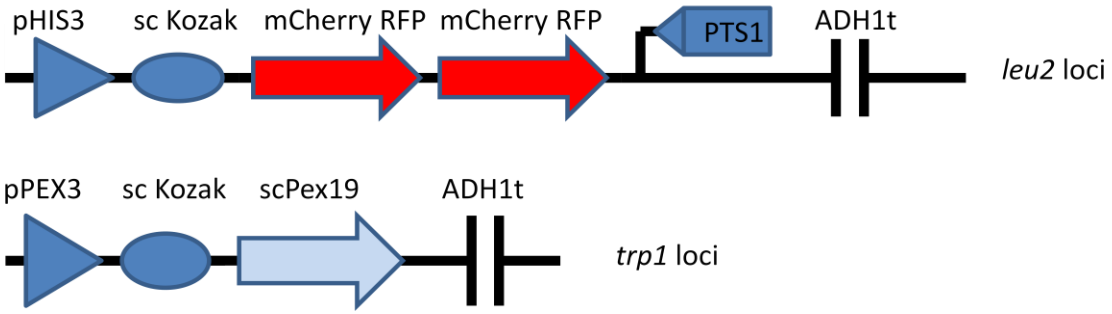


A) **B)**

Figure 5.21: (BY4741 *MAT*a, *pex3-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP) A) RFP-PTS1 peroxisomes (mCherry) B) KJY4 showing yeast *PEX3*-CFP peroxisomes (Cerulean). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.21 shows the KJY4 rescue strain. Figure 5.21 A shows RFP peroxisomes, indicating that peroxisome import function has been rescued. Figure 5.21 B shows CFP peroxisomes, with a near total overlap with the RFP ones in Figure 5.21 A. This indicates the construct has rescued *PEX3* function. It also indicates that all peroxisomes are tagged with our yeast *PEX3*-CFP construct.

5.7.5 Strain 5: KJY7 – Yeast Pex19p Rescue Control



The KJY7 strain uses the single *PEX19* knockout to test the yeast *PEX19* rescue. The yeast *PEX19* construct rescues peroxisome function. The RFP-PTS1 construct demonstrates peroxisome import function. As this is a *PEX19* rescue experiment, the original *PEX3* was not deleted.

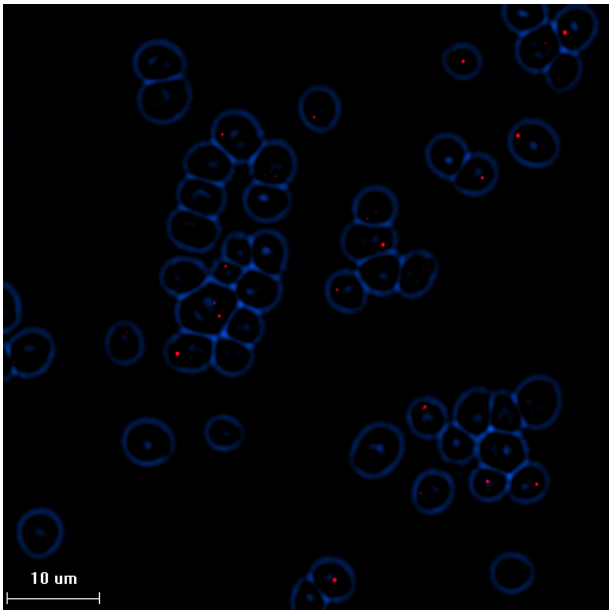


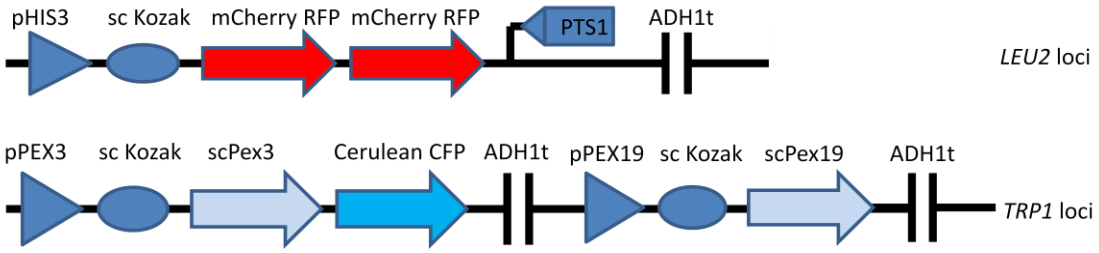
Figure 5.22: (BY4741 *MAT*a, *pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1 TRP1::pPEX19-scPEX19) RFP-PTS1 peroxisomes (mCherry). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

As indicated in the introduction to this section, Pex19p is untagged as it is a cytoplasmic peroxin (peroxisomal protein). There is only one fluorophore in KJY7 which is shown as the RFP-PTS1

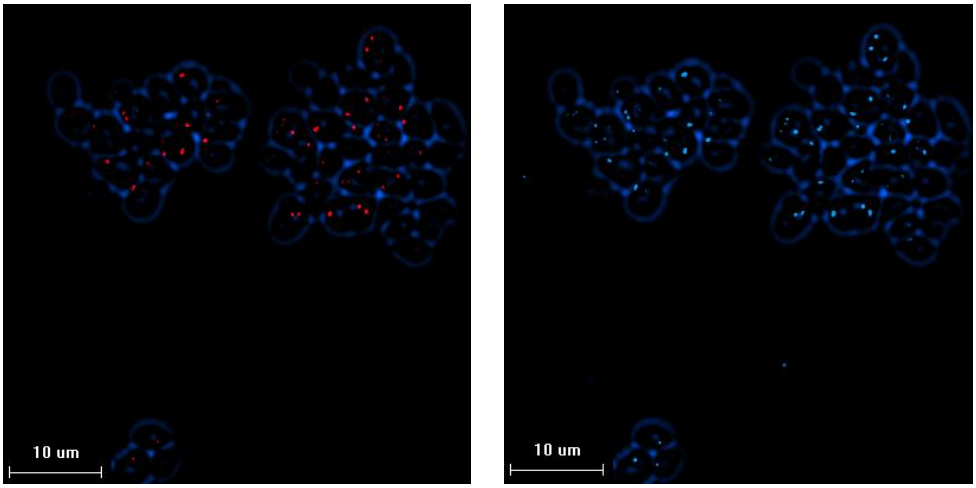
construct in Figure 5.22. Peroxisomes are visible, but there are clearly fewer than have been previously observed in the wild type and *PEX3* rescue. This version of *PEX19* may not be expressed at the correct level, but the yeast can still produce functional peroxisomes. This effect could be a function of the URA3 locus, but no further testing was conducted to verify this observation.

5.7.6 Strain 6: KJY8 – Yeast Pex3p and Pex19p Rescue Control

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex19-Δ*



The KJY8 strain was constructed to test the rescue of the double *PEX3* and *PEX19* knockout. It was transformed with two constructs, the RFP-PTS1 peroxisome function control and the yeast *PEX3*-CFP and *PEX19*. Both *PEX3* and *PEX19* were built onto the same construct to save on the number of sequential transformations needed with each strain.

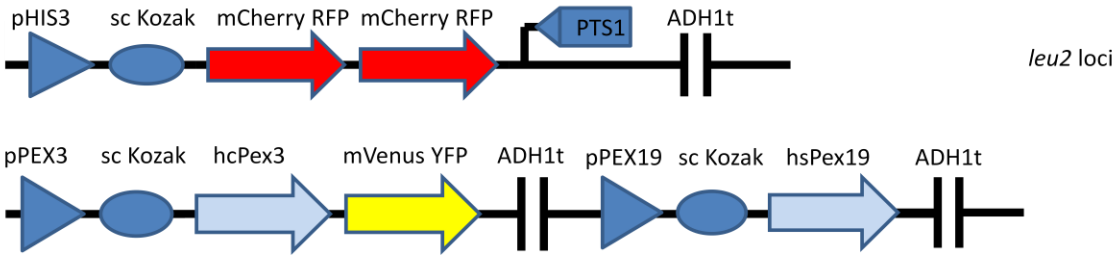


A) **B)**

Figure 5.23: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t -pPEX19-scPEX19) A) RFP-PTS1 peroxisomes (mCherry) B) scPex3p-CFP (Cerulean) peroxisomes rescued by the yeast *PEX3* and Pex19p construct. The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

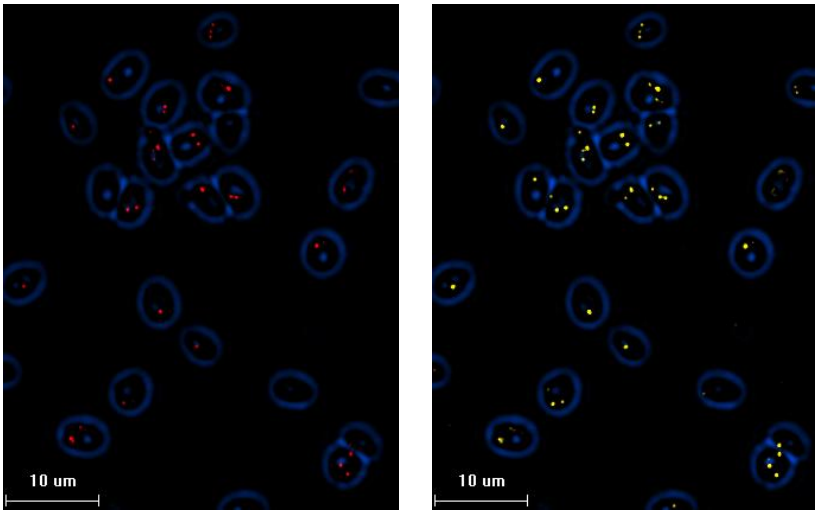
Figure 5.23 A shows functional peroxisomes with the RFP-PTS1 construct. Figure 5.23 B shows CFP peroxisomes that have been rescued with the yeast Pex3p-CFP and Pex19p constructs. These images demonstrate the successful rescue of peroxisome function when both essential genes have been removed. They further show functional CFP yeast peroxisomes in greater numbers than with the previous control in Figure 5.22.

5.7.7 Strain 7: KJY9 – Human Pex3p and Pex19p Rescue Control



The KJY9 strain was constructed to test the rescue of the double *PEX3* and *PEX19* knockout using the human genes. It was transformed with two constructs, the RFP-PTS1 control and the

human Pex3p-YFP and Pex19p. Both *PEX3* and *PEX19* were built onto the same construct to save on the number of sequential transformations.



A) **B)**

Figure 5.24: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ* LEU2::pHIS3-RFP-RFP-PTS1, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19) A) RFP-PTS1 peroxisomes (mCherry) B) KJY9 hsPex3p-YFP (mVenus) peroxisomes rescued by the human *PEX3* and *PEX19* construct. The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.24 A shows functional peroxisomes with the RFP-PTS1 construct. Figure 5.24 B shows YFP peroxisomes that were rescued with the human Pex3p-YFP and Pex19p constructs. These images demonstrate the successful rescue of peroxisome function using genes from another species when both essential genes have been removed. This represents a novel experiment and has not been performed in yeast using human genes before.

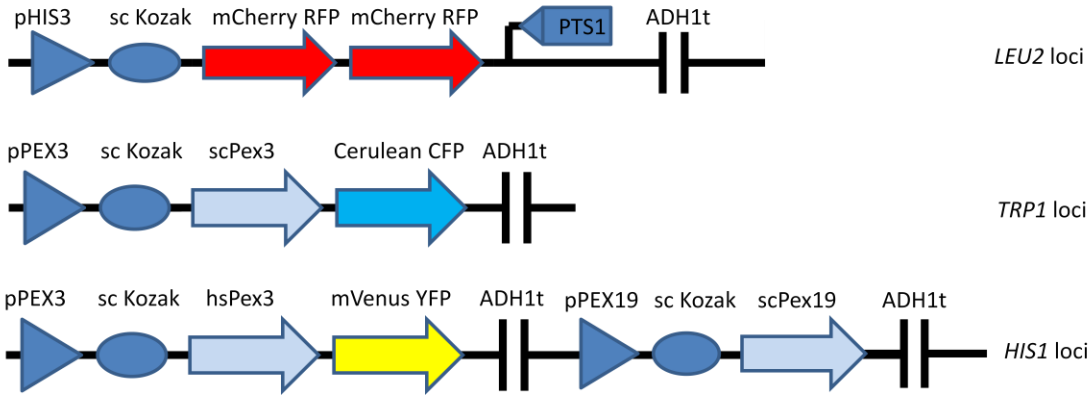
5.8 CFP-YFP Controls

The previous control experiments demonstrated successfully tagged peroxisomes with the RFP-PTS1 construct. This construct is essential to demonstrate functionality, as the peroxisome import complex must be assembled and functional for import to be successful. The second category of controls demonstrated successful deletion of yeast *PEX3* and *PEX19*. Further controls showed the ability to rescue both single knockouts and the *PEX3* and *PEX19* double

knockout. Both human and yeast functional peroxisomes were shown, allowing the project to progress to the next stage in development. Furthermore, peroxisome function was rescued using human Pex3p and Pex19p, which has not been described in previous literature. This section illustrates the three tri-channel fluorophore controls before showing the main experimental strain.

5.8.1 Strain 9: KJY11

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ, pex19-Δ*



KJY11 is the same as the main experimental strain minus the yeast *PEX19*. This control aims to artificially create two populations of peroxisomes by introducing a bias in the way they are formed.

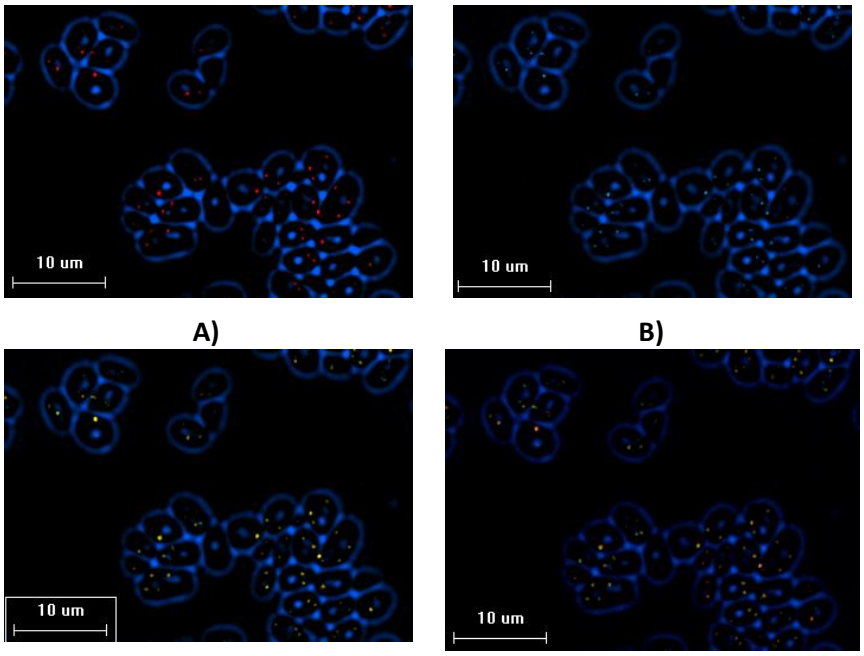
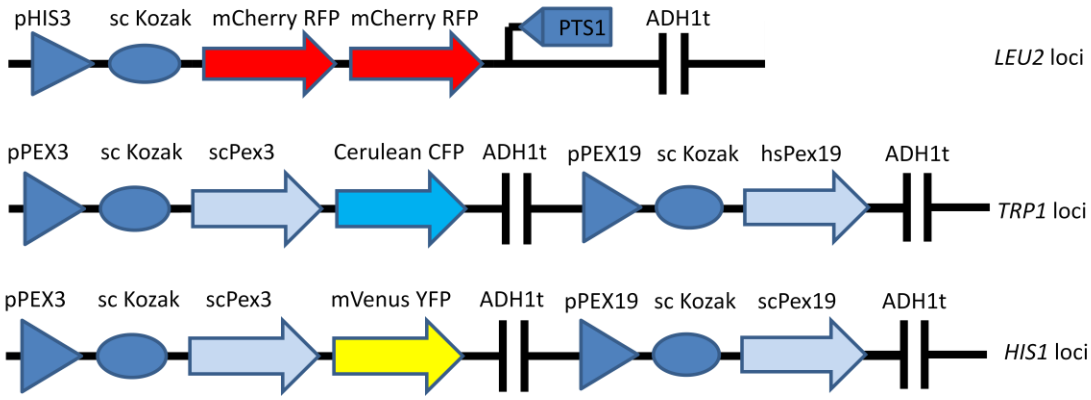


Figure 5.25: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19) A) RFP-PTS1 peroxisomes (mCherry) B) scPex3p-CFP peroxisomes (Cerulean) C) hsPex3p-YFP peroxisomes (mVenus) D) yeast and human false color overlay (Cerulean and mVenus) (CFP = green, YFP = red). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.25 A shows functional peroxisomes with the RFP-PTS1 construct. Figure 5.25 B shows the yeast Pex3p-CFP construct. Figure 5.25 C shows the human Pex3p-YFP peroxisomes. Figure 5.25 D is a false color CFP-YFP image where CFP is green and YFP red. The KJY11 control was constructed to create a bias in the peroxisome populations be removing the yeast shuttle protein, Pex19p. There is no observed difference in the CFP and YFP channels in B and C. There is little color difference across the population in the false color overlay in D, a point that will be demonstrated numerically in the next chapter.

5.8.2 Strain 10: KJY14

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ, pex19-Δ*



KJY14 is a control strain where both CFP and YFP fluorophores are tagged to the yeast *PEX3*. Both yeast and human copies of *PEX19* are present. The aim of this control was to create two populations of peroxisomes with a bias on the yeast population due to the lack of the human tagged Pex3p.

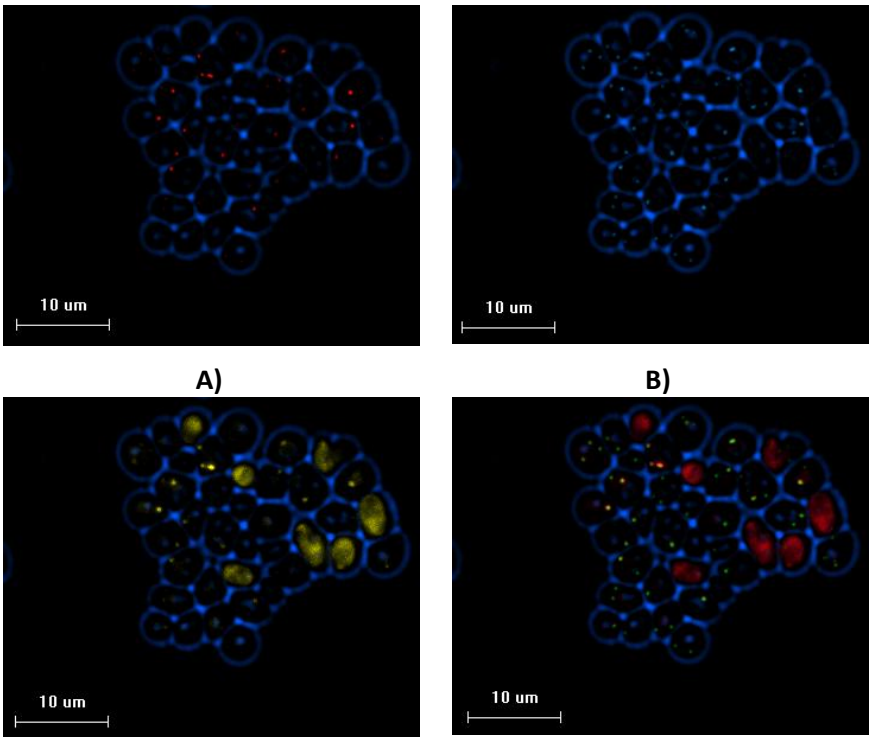


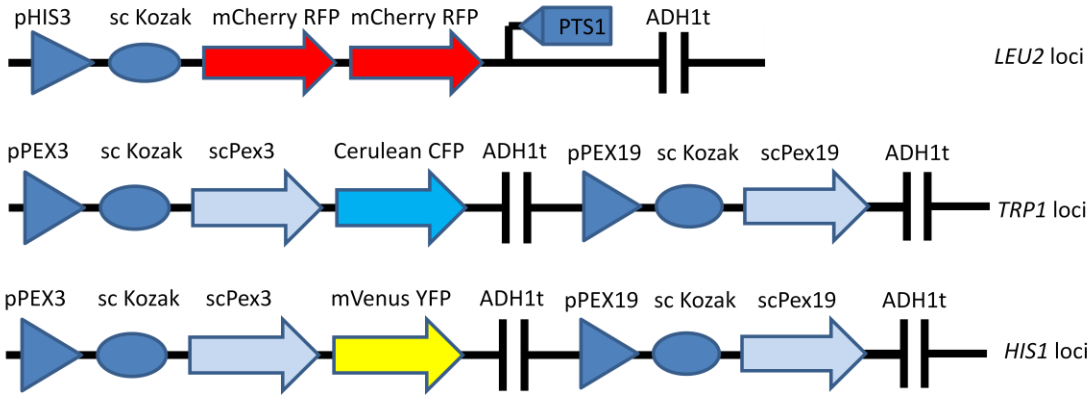
Figure 5.26: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-hsPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-scPEX19) A) KJY14 RFP-PTS1 peroxisomes (mCherry) B) scPex3p-CFP peroxisomes (Cerulean) C) scPex3p-YFP peroxisomes (mVenus) D) Pex3p false color overlay (Cerulean and mVenus) (CFP = green, YFP = red). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.26 A shows functional peroxisomes with the RFP-PTS1 construct. Figure 5.26 B shows correctly formed yeast Pex3p-CFP peroxisomes. Figure 5.26 C shows some yeast Pex3p-YFP peroxisomes in some cells, but cytoplasmic localization in a small number of the yeast cells. In the defective cells, no yeast-YFP peroxisomes are apparent. This could be from Pex3p-YFP that remains bound to human Pex19p. Part of the problem could also be a peroxisome inheritance issue between mother and daughter cells. If the population is not homogenous and some of the mother cells do not form peroxisomes and pass this phenotype to their daughter cells, sporadic cytoplasmic localization of Pex3-YFP could occur. Figure 5.26 D shows a false color CFP-YFP image where CFP is green and YFP red. There is little overlap between the two colors and the

cytoplasmic localization will interfere with the numerical analysis, as the problem is not present on the RFP channel.

5.8.3 Strain 1: KJY15

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ, pex19-Δ*



KJY15 is a control strain where both CFP and YFP fluorophores are driven by the yeast *PEX3*. This control aims to create a single, uniform population of mixed CFP and YFP peroxisomes as there is no copy of the human Pex19p.

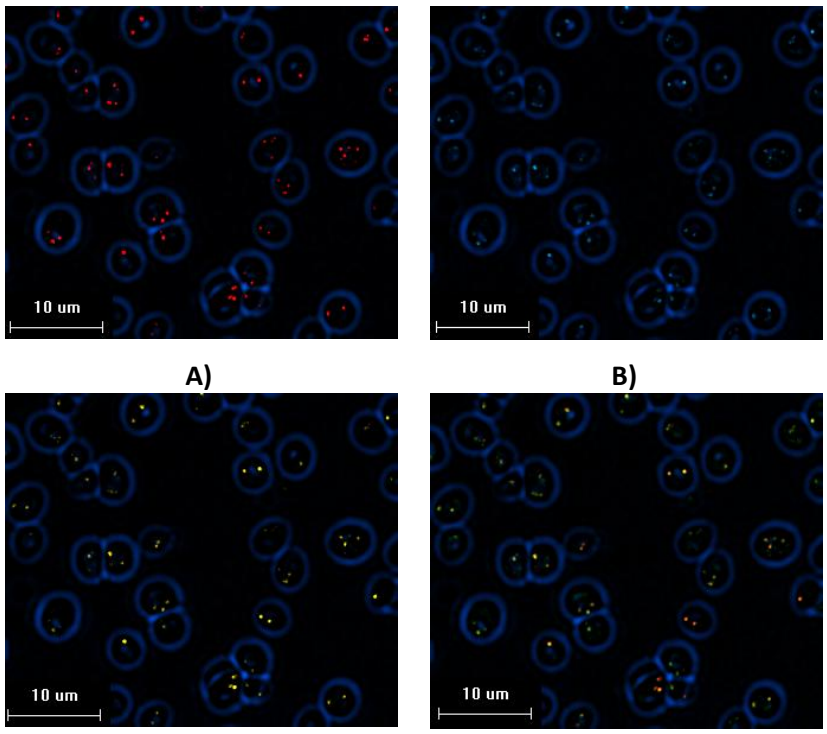


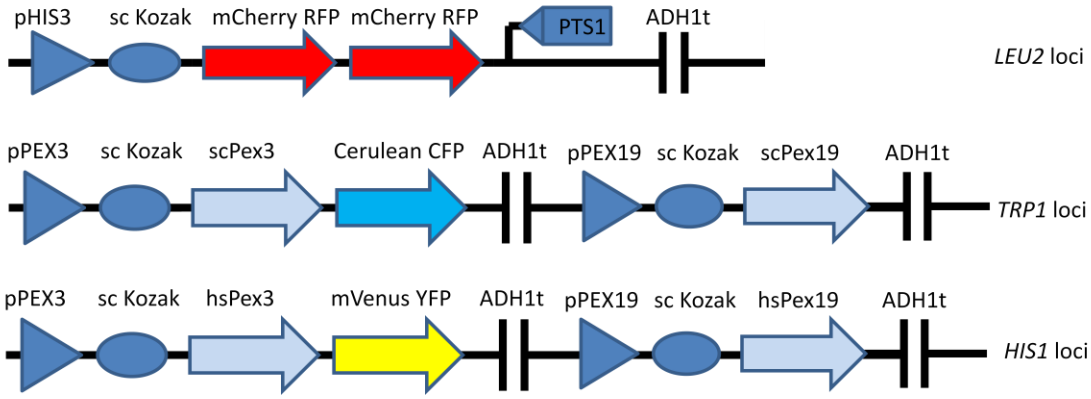
Figure 5.27: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-scPEX3-YFP-ADH1t-pPEX19-scPEX19) A) RFP-PTS1 peroxisomes (mCherry) B) scPex3p-CFP peroxisomes (Cerulean) C) scPex3p-YFP peroxisomes (mVenus) D) yeast and yeast false color overlay (Cerulean and mVenus) (CFP = green, YFP = red). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.27 A shows functional peroxisomes with the RFP-PTS1 construct. Figure 5.27 B shows yeast Pex3p-CFP peroxisomes. Figure 5.27 C shows yeast Pex3p-YFP peroxisomes. Figure 5.27 D is a false color CFP-YFP image where CFP is green and YFP red. The CFP and YFP channels in B and C appear to be closely correlated, but the peroxisomes are not a uniform yellow in D (which would represent a homogenous population of peroxisomes). There is some difference in the CFP and YFP populations, but the images do not represent clear bimodality. The numerical analysis in chapter 7 describes the nature of this peroxisome population, which will be shown to be heterogeneous.

5.9 Main Experimental Strain

5.9.1 Strain 8: KJY10

BY4741a genotype: *MATa, ura3-Δ0, leu2-Δ0, met15-Δ0, his3-Δ1, pex3-Δ, pex19-Δ*



The KJY10 strain is the main experimental strain for the project. It has been transformed with three constructs. The first is the RFP-PTS1 to test for peroxisome import function. The second is the yeast Pex3p-CFP and Pex19p construct, to induce yeast-CFP peroxisomes. The third is the human Pex3p-YFP and Pex19p construct to induce human peroxisomes.

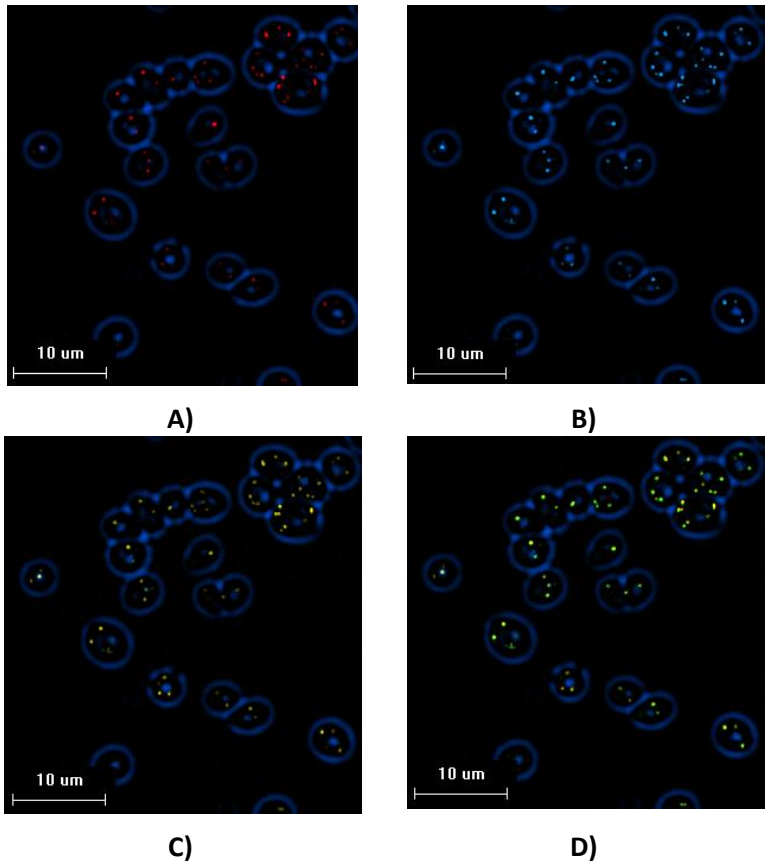


Figure 5.28: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) A) RFP-PTS1 (mCherry) peroxisomes, B) scPex3p-CFP peroxisomes (Cerulean) C) hsPex3p-YFP peroxisomes (mVenus) D) yeast and human false color overlay (mVenus and Cerulean) (CFP = green, YFP = red). The strain was grown in YPDA for 24 hours then washed in ddH_2O and transferred to synthetic raffinose media to induce peroxisomes. The cells were imaged fixed within 48 hours of fixation.

Figure 5.28 A shows the RFP-PTS1 construct and functional peroxisomes. Figure 5.28 B shows the yeast Pex3p-CFP peroxisomes. Figure 5.28 C shows the human Pex3p-YFP peroxisomes. Figure 5.28 D is the first false color composite CFP-YFP image. As explained in the introduction to this Chapter, CFP is represented by green and YFP by red. Yellow represents an overlap of the two colors. These images show most of the peroxisomes appearing as yellow/light green indicating that there is little evidence of two species specific populations of

peroxisomes. Numerical analysis of these images in the following chapter shows the population of peroxisomes is a normal distribution.

5.10 Conclusions

When imaging, the acquisition times and brightness settings were calibrated based on the fluorophore. CFP was generally imaged for longer than YFP and once the settings were chosen, up to 10 images were imaged using identical settings. Z stacks were taken of the fluorescent channels, bright-field cells and a blank region on the slide to perform a background subtract. Once acquired, the Z-stacks were 3D deconvolved using AutoQuant X and processed using a script written for Image Pro 7. This script identifies the peroxisomes and produces an Excel spreadsheet containing the CFP, YFP and RFP peroxisome data.

Initial microscopy was performed on the KDY3-KX9-KX10-KX9_URA3 strain. This was achieved and the dual color overlays indicate a bimodal population. However, it was later discovered the xenon bulb was damaged and created an artificial bias in the image. This image was discarded and microscopy was performed on the BY4741a strain collection. The control strains were received, grown and imaged with the protocols described above. The rescue controls all proved successful and neither the problem with cytoplasmic fluorescence nor the malfunctioning RFP-PTS1 were observed. The new strain was deemed fit for the project and imaging could proceed with the key controls and main experimental strain.

KJY11 was imaged and demonstrated both CFP and YFP peroxisomes. KJY14 showed a problem as some peroxisomes were visible, but a lot of cytoplasmic localization was also observed. KJY15 was imaged and visually showed a good CFP and YFP peroxisome overlap. Finally, KJY10 was successfully imaged and analyzed. However, it did not show much visual difference between the CFP and YFP channels. Numerical analysis of the results produced using the Image Pro analysis script is necessary to determine correlation between CFP and YFP for the three key controls and the main experimental strain. This analysis will be conducted in the following Chapter.

6 Numerical Analysis of the Peroxisome Experiments

Images of the YPH500α and BY4741a strains were presented in the previous Chapter. This Chapter describes methods used to collect and analyze the quantitative data captured from these images. Data from the PSY1198-α strain are shown using the initial two channel Matlab processing script to highlight some of the issues with this analytical method. The BY4741a strain analysis was conducted using a three channel normalization protocol based on functional peroxisome information in the RFP channel.

Once the peroxisome data was captured with image manipulation scripts written in Image Pro 7.0, it was analyzed in a quantitative manner. The peroxisome average intensity was recorded for each peroxisome and used for the analysis. This is a measure of the intensity of each pixel, averaged over the area of the peroxisome. Before the images were analyzed, this average intensity was examined on an image to image basis using a simple moving average.

6.1 Contributions to this Chapter

The basis of the first Matlab script was written by Jake Wintermute. The author wrote scripts for the more comprehensive two channel analysis and normalization and further three channel normalization and statistical analysis. David Kelly wrote the initial image analysis scripts in Image Pro 7 according a protocol that was determined by the author. The author made all further modifications and alterations to the Image Pro 7.0 analysis scripts.

6.2 Matlab Processing

The first Matlab script was initially written to visualize the relative intensity of CFP and YFP peroxisomes in a histogram as recorded by MetaMorph. The code served as a basis for versions that were completely re-written by the author. It was made as a simple way of viewing the CFP to YFP ratio. The script starts by loading in the Excel data file from MetaMorph and assigning a channel to the CFP and YFP data. This channel data is then parsed into a histogram by dividing CFP by YFP and plotting the log of the ratio. A moving average of the CFP and YFP raw data were also plotted to examine the image to image variation across the data-set. As each image

has been individually thresholded by eye from the MetaMorph processing, examining this parameter is important for the integrity of the data. The peroxisome number scale represents the intensity of every peroxisome in the series plotted sequentially and the intensity is a dimensionless measure of the grayscale value measured during the experiment.

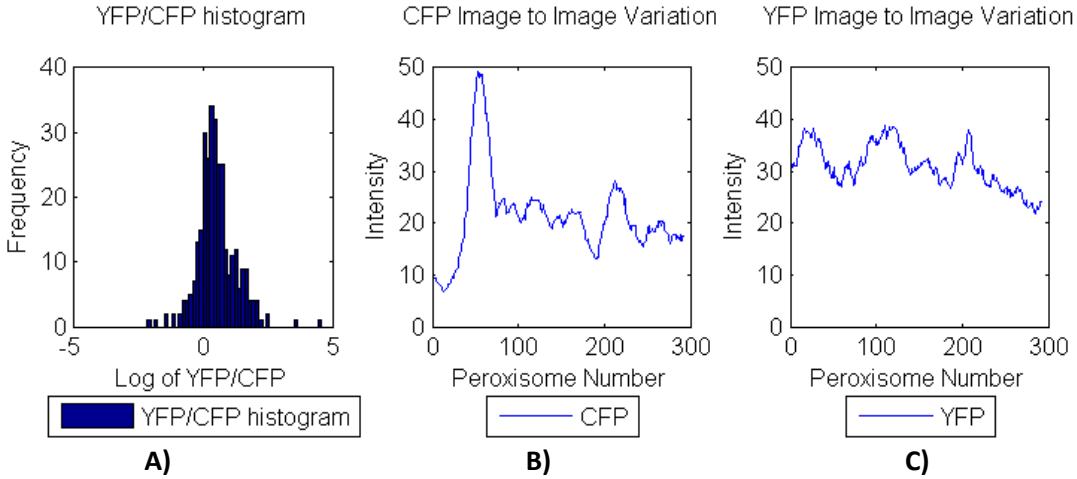
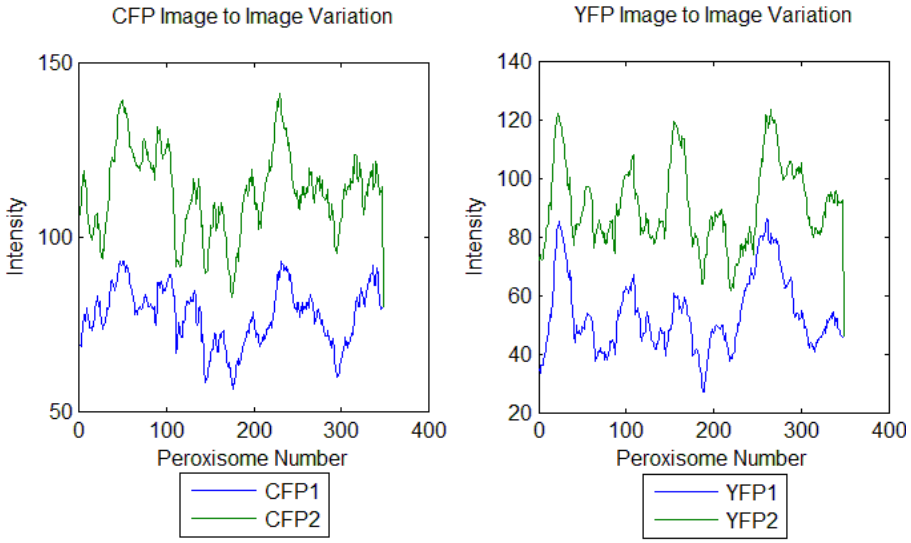


Figure 6.1: Matlab plots from the first *pex* processing script. A) Histogram of the YFP/CFP ratio against the frequency of the peroxisomes B) CFP image to image variation in peroxisome average intensity C)YFP image to image variation in peroxisome average intensity.

Figure 6.1 shows the plots from the original Matlab script describing YPH500α. The strain was grown to mid log phase in YEPD, then washed and transferred into YPGO with added adenine for 24 h. The cells were paraformaldehyde fixed and imaged suspended in mounting media. The images were produced by acquiring a single best focus plane with a 2 second CFP exposure and a 1 second YFP exposure. They were then analyzed using the MetaMorph protocol described in the previous section.

Figure 6.1 A shows a histogram of the log of the YFP to CFP ratio. This ratio gives an indication of how much CFP a peroxisome has relative to the amount of YFP. In this case, the plot looks roughly like a normal distribution. Figure 6.1 B and C show a moving average of the peroxisome average intensity. The average intensity of each peroxisome is plotted in sequence. This parameter is necessary to visualize at an early stage as significant image to image fluctuations could be problematic when analyzing a ratio. The purpose of these plots was to get a quick idea of the CFP/YFP ratio and to examine the image to image variation. This was to test

the image and data analysis and determine if and what changes were needed. An in depth analysis was not conducted until much later in the project.



A) **B)**

Figure 6.2: Photo-bleaching control A) CFP (Cerulean) image to image average intensity data. B) YFP (mVenus) image to image average intensity data. Green represents the first exposure while blue shows a second exposure of equal intensity and duration five seconds later Note: the y-axis have not been normalized to highlight the relative difference between channels.

Figure 6.2 shows the detail in the difference between the first and second images acquired during the photo bleaching experiment. There is a shift in the scale of the intensity, but not the YFP/CFP ratio. Photo bleaching was therefore not considered to be a significant hurdle in the progression of the project.

6.3 Image Analysis Macro Evolution

Several versions of the image analysis macro were created as the script evolved. Version 4 included an additional peroxisome identification step on the CFP and YFP channels. This step creates duplicates of the CFP and YFP single plane images and applies the same image manipulation processes as performed on the RFP channel. This process introduces an image to image bias due to the manual thresholding now being performed on all three channels.

Macro V5 is an elaboration of V2 where peroxisome data are recorded for the RFP channel along with the CFP and YFP data. This information was not recorded in previous versions as it was not deemed necessary. The RFP regions were used to identify functional peroxisomes and these regions were reported from the CFP and YFP data. Recording the additional RFP data allows a more thorough analysis to be conducted. Initially, this RFP peroxisome data were to be used to normalize the CFP and YFP intensities on a peroxisome to peroxisome basis. When this individual normalization was attempted, it removed much of the CFP/YFP resolution which was not acceptable. The RFP channel will be used to normalize on an image to image basis. A summary of the version progression can be seen in Table 6.1.

Version	Functions
V1	Dual channel; only plots peroxisome data.
V2	Dual channel image to image and population normalization.
V4	Tri channel, image to image and population normalization, reports CFP and YFP peroxisome data.
V5	Tri channel, image to image and population normalization, reports CFP, YFP and RFP peroxisome data.

Table 6.1: Matlab analysis script progression

6.4 Dual Channel Data Normalization

Once the images were acquired and the data were processed using the ImPro analysis script, it was imported into Matlab for numerical analysis and normalization. Several parameters were measured including: peroxisome area (per peroxisome, not averaged over the cell), average intensity per peroxisome, maximum intensity per peroxisome, the peroxisome origin and the channel (CFP, YFP or RFP). The average intensity was chosen as the most important representation of the information contained within each peroxisome region and a script was written to normalize the data based on this parameter. The CFP and YFP channels were normalized to each other based on the image to image average intensity and then the sample average.

6.5 Data Acquisition using AutoQuant X and Image Pro 7

Images for this project were processed using a set of software including Image Pro V7.0 (ImPro), AutoQuant X and Matlab. The images were captured in .dv format on the microscope and converted to .seq using ImPro. They were then opened in AutoQuant X to perform a 3D

deconvolution. The .seq files contain all the information necessary for the deconvolution process (sensor pixel size, z-stack spacing, etc...) except for the objective magnification and the immersion media. The images are always captured using a 100x objective with oil immersion and this information needs to be manually entered into the processing software.

6.5.1 Image Pro analysis Macro

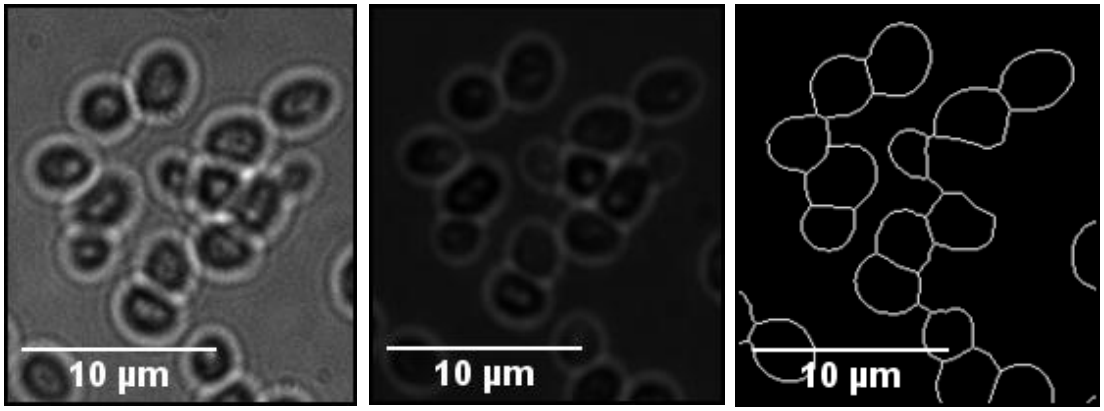
The image analysis macro is presented in this Chapter. A copy of this script is appended on page 197, Appendix 7: Image Pro 7.0 Macro V.5 script. Once the images were deconvolved, they were processed to extract peroxisome data using a macro in ImPro. Several versions of this script were developed as the analysis requirements evolved.

The purpose of this script was parse out the peroxisome data from qualitative images into a quantitative format that can be numerically analyzed. This was done with an emphasis on adding as little bias to the CFP/YFP ratio as possible. It has a function to track each peroxisome to its cell of origin and will not measure any peroxisome intensity readings outside a cell. This helps to cut down on false positive peroxisomes by eliminating any fluorescent contamination or minor image artifacts from the acquisition and deconvolution process. The script can be broken down into three main modules:

1- Creation of cell regions from the out-of-focus bright-field z-stack.
2- Manipulation of the RFP image to identify user-defined peroxisome regions then reapplication of those regions to the CFP, YFP and unmanipulated RFP images.
3- Excel spreadsheet output of peroxisome quantitative data.

6.5.2 Image Pro Script Module 1 Details

The first module of the script begins by subtracting a nominal numerical value from the blank z-stack image. Subtracting this value is essentially a trick to increase contrast on the important regions of interest in the next processing step. To produce the blank, a region of the slide with no cells is imaged using the out-of-focus bright-field.



A) **B)** **C)**

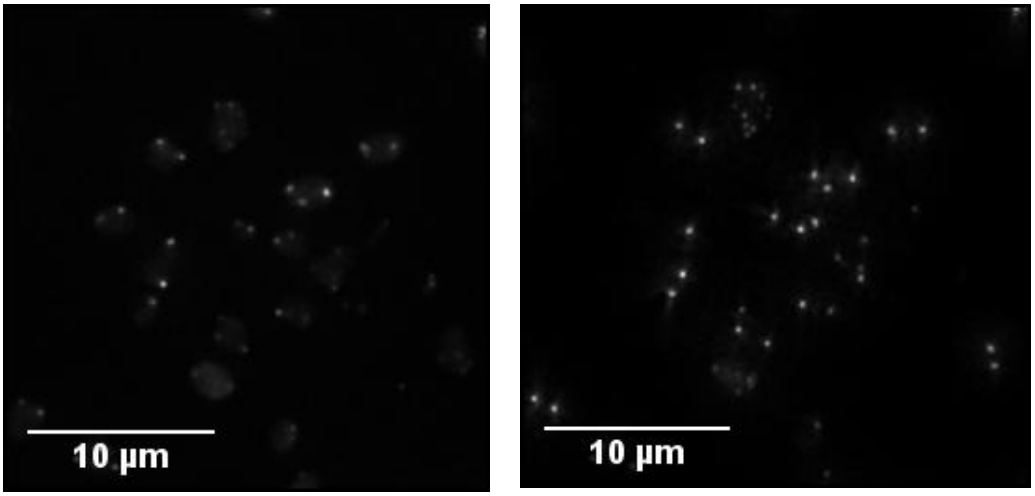
Figure 6.3: (BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) out-of-focus bright-field z-stack B) out-of-focus bright-field z-stack after background subtract C) final cell regions after skeletonizing process. Cells were imaged fixed.

This is microscopy best practice as this blank can then be subtracted from the regions that contain cells as shown in Figure 6.3 A to remove any irregularities. This blank is subtracted from the out-of-focus bright-field image containing the cells. The result of this subtraction is a z-stack that contains half dark cells and half bright "halos" Figure 6.3 B. The halo regions are selected for further processing and the other parts of the stack are deleted. The halo stack (usually 10 – 20 images) is compressed into one frame and a smoothing filter is applied. This single frame then has a sharpening "large spectral filter" applied to increase the contrast in the regions defined as the cell outlines. Finally, a skeletonizing filter is run to create cell boundaries that are 1 pixel thick (Figure 6.3 C). It is important to note that this processing technique does not give the exact size of the cells. It also does not count 100% of the cells, as some of the cells present in Figure 6.3 A do not appear in C. Algorithms to distinguish mother from bud and to ignore clumps several cells deep were not implemented. The analysis can easily identify single cells but cannot always pick up grouped cells due to the smaller relative difference in contrast between the cell boundaries. It gives a good indication of how many peroxisomes are in each cell and which peroxisomes come from which cell, but the areas defined as cell regions are larger than the actual cells are joined in regions of high cell density. This effect can be seen in the control strains in Figure 5.19 and Figure 5.20 where the cytoplasmic localization of the cells does not exactly match up with the cell regions.

In order to create regions around the cells, an embedded script is run within the first module. To illustrate the non-trivial nature of the analysis, this script was written by the ImPro software developers specifically for this analytical situation. The actual type of analysis is known as "fried egg analysis" as the program seeks to identify yolks (peroxisomes) within egg whites (cells) within a frying pan (the whole image). The code uses the skeletonized cell regions to identify regions that are size selected for yeast cells. As this technique only identifies regions of the correct size as cells, no fluorescent contamination appearing outside cells is falsely identified as a peroxisome.

6.5.3 Image Pro Script Module 2 Details

Once the first module was run, the second begins by creating a single image of the 50 images that were captured in a z-stack in the CFP, YFP and RFP channels as shown in Figure 6.4.

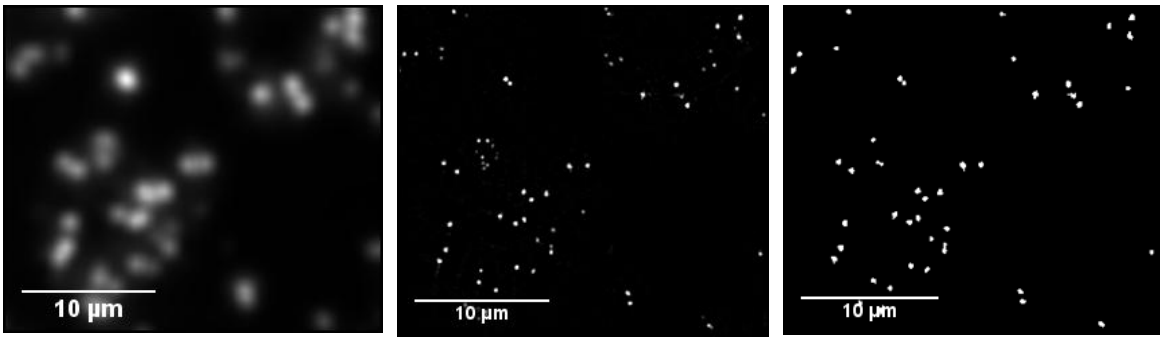


A) **B)**

Figure 6.4: (BY4741 *MAT*a, pex3-Δ, pex19-Δ, LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) Single plane of the 3d deconvolved compressed Z-stack showing YFP peroxisomes (mVenus) B) Maximum brightness extended depth of field YFP image (mVenus). Cells were imaged fixed.

A maximum brightness "extended depth of field" operation is run to compress the maximum intensity pixels from the ~50 images into the same plane (Figure 6.4 B). The CFP and YFP single plane images are not manipulated any further. Any further modifications will result in variations to the YFP/CFP scaling information which could introduce bias.

All further manipulations are performed on the RFP channel to identify the regions deemed peroxisomes in order to reapply and measure those areas on the CFP, YFP and original RFP images. Using the RFP channel is necessary to prevent introducing bias on the CFP and YFP channels. The RFP channel is used as these peroxisomes have been identified as functional by their import of the RFP-PTS1 marker. Those functional peroxisome regions can then be applied back to the CFP, YFP and unmanipulated RFP image to get quantitative data. Two copies of this RFP maximum intensity image are created and one is subject to a blurring large spectral filter operation to smooth out any fluorescence information that might not come from the peroxisomes.



A) **B)** **C)**

Figure 6.5: (BY4741 MATa, pex3-Δ, pex19-Δ, LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) A) Large spectral filter blurred image, B) Large spectral filter sharpened image, C) Binary mask. Cells were imaged fixed.

The blurred filtered image (Figure 6.5 A) is subtracted from the max intensity image to sharpen the peroxisomes. A second sharpening large spectral filter is then run on this image to increase the contrast in the peroxisome regions as shown in Figure 6.5 B. The image is manually thresholded to correspond to the expected peroxisome regions. This technique does introduce some bias, but this bias exists only on an image to image basis. These regions are transformed into a binary mask (Figure 6.5 C). This binary mask is then inverted and subtracted from the CFP, YFP and original RFP images. The resultant images only have signal on regions that the technique has identified as peroxisomes and everywhere else on the image has a zero value.

Size filters are set to correspond with the maximum and minimum expected peroxisome size and the final module can then be applied.

6.5.4 Image Pro Script Module 3 Details

The last module in this macro takes information about the cell regions and creates data from the areas identified as peroxisomes, starting with the YFP image. Each cell is individually called up and the intensity is examined based on the peroxisome image created from the RFP mask. The size is applied to sort out some of the background signal. Once a region has been identified as being a peroxisome of the correct size within a cell, its area, roundness, average brightness, maximum brightness and the cell number that it came from are written to an Excel spreadsheet. This process is then repeated for the CFP and RFP images. This spreadsheet is manipulated in Excel or read into Matlab for analysis.

Fluorophore code	KJY15_1 CFP Area	Max Intensity	Average Intensity	Cell Number
2	2.42440	49611	23932.68	1
2	2.04160	40558	23399.25	1
2	2.04160	22080	13813.81	1
2	1.53120	22706	15000.92	3
2	1.40360	13090	9667.091	3
2	2.04160	29420	16665.06	5
2	1.53120	34499	21447.75	6
2	1.53120	39999	24039.83	6

Table 6.2: Example spreadsheet from peroxisome macro

Table 6.2 shows an example of the spreadsheet generated from the ImPro process. The fluorophore code identifies the peroxisome as RFP, CFP or YFP. The area is the area (in um^2) of the peroxisome. Max intensity is the brightest pixel in the area and the average is calculated from all pixels in the peroxisome. Finally, cell number tracks the cell of origin of the peroxisome.

6.5.5 Image Pro Image Processing Summary Table

Table 6.3 shows a summary table of the Image Pro processing script. This is a summary of the techniques described in this Chapter on how the images are analysed and parsed for data capture.

DeltaVision – Image Pro 7.0	**Module:**
Run 3d deconvolution	-
Load out-of-focus brightfield Z-stack	1
Subtract nominal value from out brightfield background stack	1
Compress half of stack into single focal plane	1
Use skeletonizing filter to select cell outlines	1
Load CFP, YFP and RFP Z-stacks	2
Compress stacks to single images using a maximum intensity projection	2
Identify peroxisomes on RFP channel	2
Threshold peroxisome regions on RFP channel	2
Create and invert binary mask on RFP channel	2
Subtract inverted binary mask from CFP and YFP channels	2
Read in peroxisome information in on CFP and YFP channels	3
Parse data into Excel spreadsheet	3

Table 6.3: MM vs. ImPro image analysis comparison table.

6.6 CFP-YFP Normalization and Image to Image Average Intensity

This script was adequate for a first pass examination of the YFP/CFP ratio and a brief examination of the CFP and YFP image to image variation. Due to the image to image variation, a normalization protocol was also included in the analysis. This script starts by reading in the dual channel CFP-YFP data. Each intensity measurement is the average of the peroxisome area measured on the CFP or YFP channel. Each data point represents a CFP or YFP measure of a peroxisome.

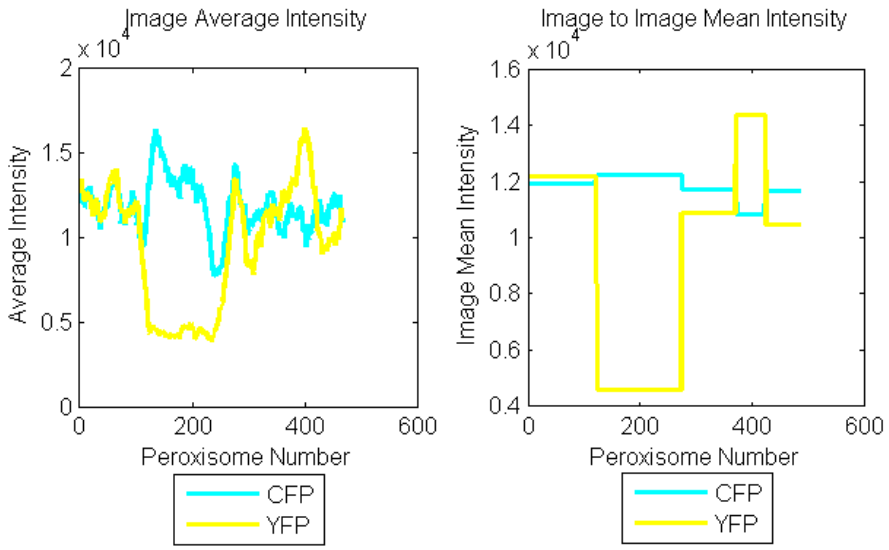


Figure 6.6: CFP (Cerulean) and YFP (mVenus) moving average and image to image average intensity. The left plot shows the raw image to image average intensity while the right plot shows the image to image average, represented by each plateau. In this data-set there are 5 images, with a clear disparity on.

Figure 6.6 shows a moving average of each peroxisome in the image variation in the first plot and the image average intensity in the second. Each plateau represents the average intensity value of an image. This representation now clearly shows that there is a significant drop in the average intensity in the second YFP image.

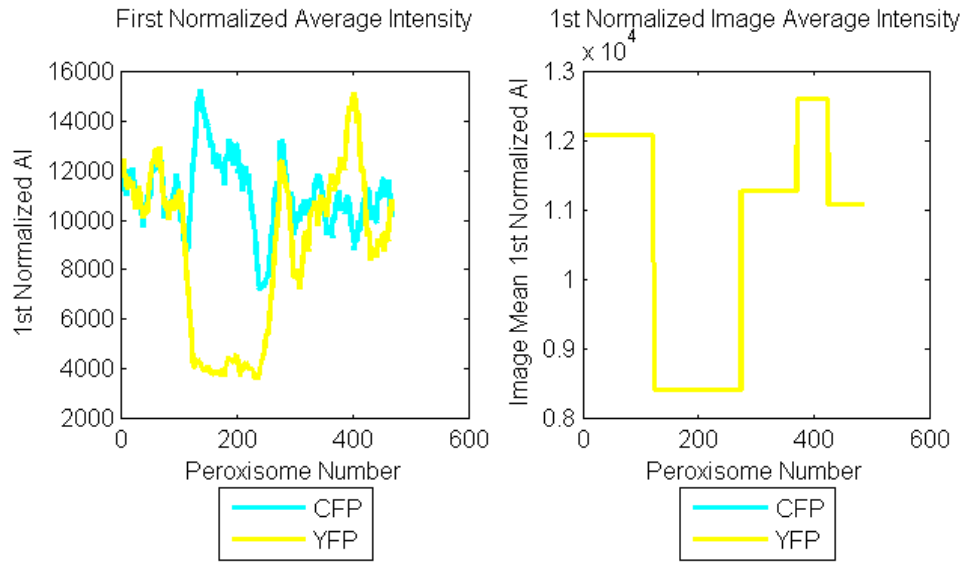


Figure 6.7: CFP (Cerulean) and YFP (mVenus) image to image normalized moving average and image to image average normalized intensity. The left plot shows the first normalized image to image average intensity while the right plot shows the first normalized image to image average, represented by each plateau. In the second plot, the CFP line is shown behind the YFP.

Figure 6.7 shows the image to image average after it has been normalized. The CFP amplitude has increased as it was offset by the relative change that was subject to both data-sets.

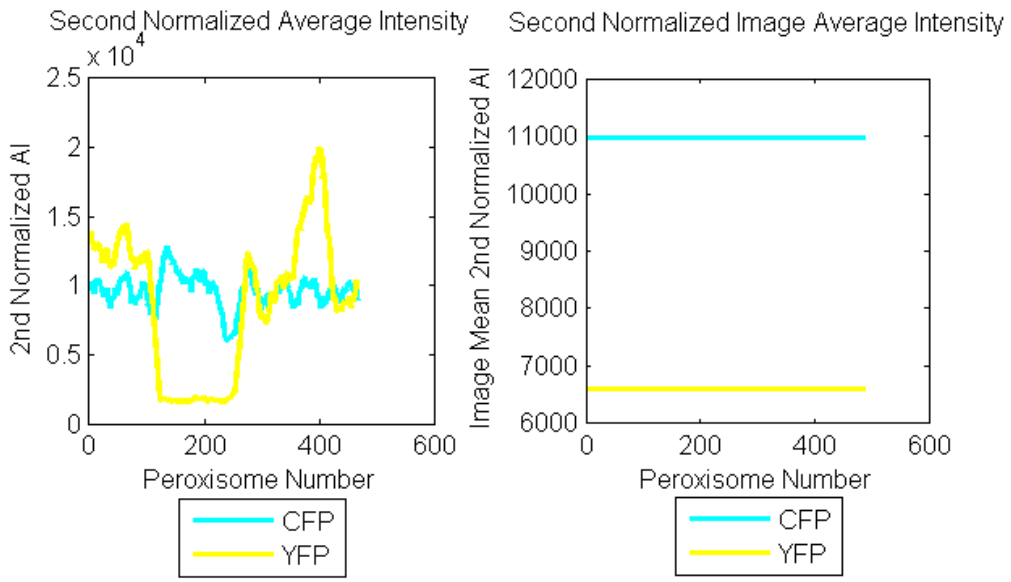


Figure 6.8: CFP (Cerulean) and YFP (mVenus) image to image second normalized moving average and image to image average second normalized intensity. The left plot shows the second normalized image to image average intensity while the right plot shows the second normalized image to image average, represented by a flat line. The second normalized plot shows two separate lines that are not superimposed.

Figure 6.8 shows an attempt to normalize the CFP and YFP data to the same average. The first plot shows the adjusted image values and the second, the image to image average. This plot should show one line representing a normalized image to image average. However, this normalization process required an iterative numerical solution to compare the CFP and YFP image to image average. This solution was not possible to achieve in the time available for analysis and was deemed unnecessary as it would introduce a bias that could affect the CFP/YFP histogram.

6.6.1 CFP-YFP Scatter Plot and Histogram Examination

Figure 6.9 shows plots of the YFP vs. CFP scatter plot and the corresponding histogram. The scatter plot shows areas of increased density in red and areas of low density, in blue.

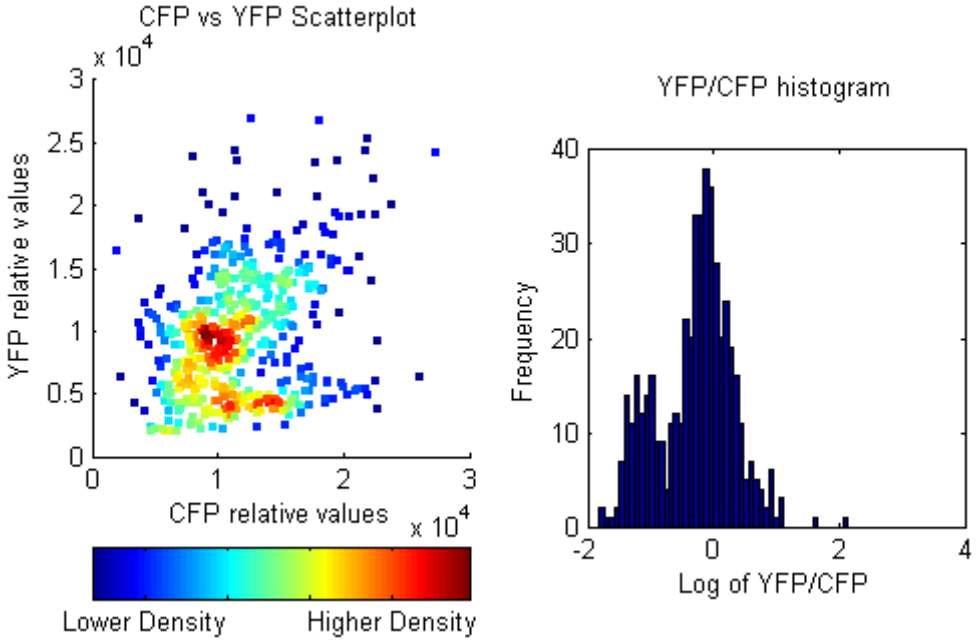


Figure 6.9: YFP (mVenus) against CFP (Cerulean) raw average intensity scatter plot and YFP/CFP peroxisome average intensity histogram.

This plot shows two areas of increased density, representing two populations of peroxisomes in the sample. The second plot in the figure shows a histogram of a log of the YFP to CFP ratio. This histogram clearly shows two populations, corroborating the information in the scatter plot.

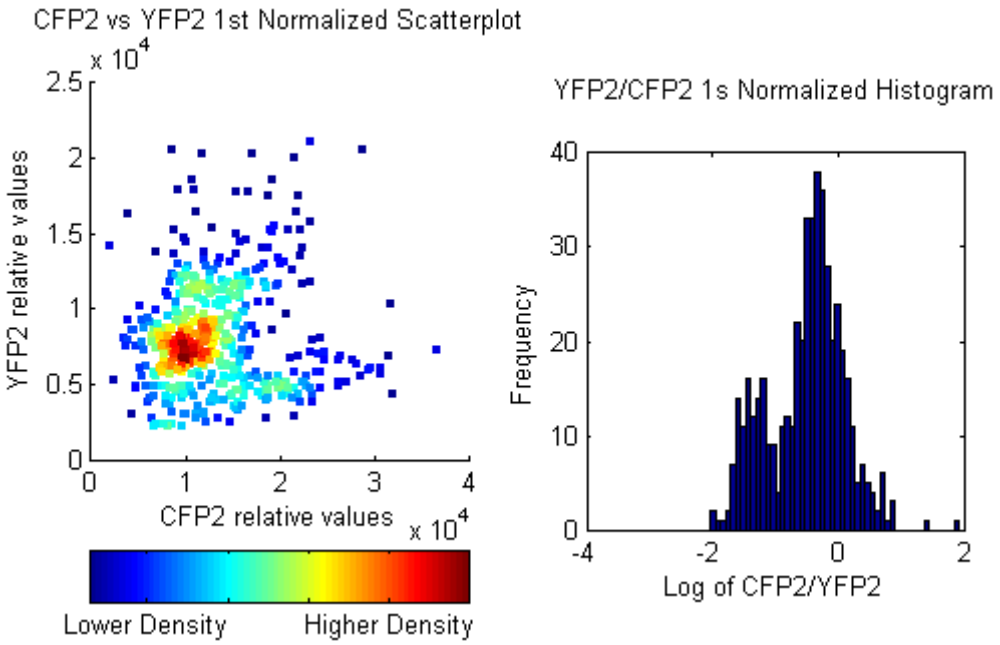


Figure 6.10: CFP (Cerulean) against YFP (mVenus) first normalized average intensity scatter plot and first normalized YFP/CFP peroxisome average intensity histogram.

The first normalization has changed the shape of the scatter plot, as there now only one area of high density is shown, relative to Figure 6.9. It does not appear to have affected the shape of the histogram.

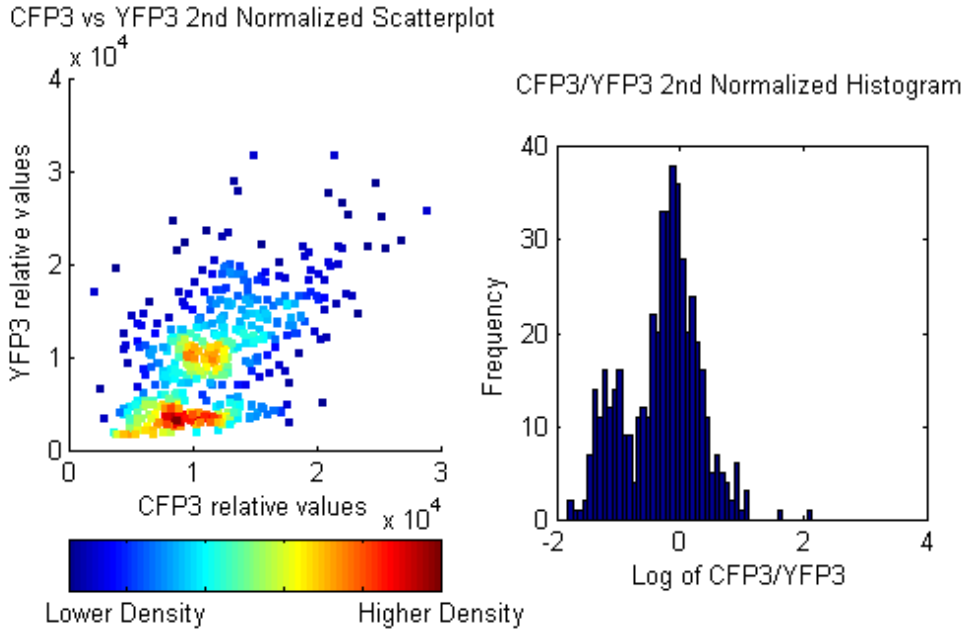


Figure 6.11: (Cerulean) against YFP (mVenus) second normalized average intensity scatter plot and second normalized YFP/CFP peroxisome average intensity histogram. The scatter plot shows two areas of high density, while the histogram shows two peaks.

Figure 6.11 shows the scatter plot and histogram for the second normalized (all image average) data. The histogram shows very little change from Figure 6.10 and Figure 6.11. However, the scatter plot shows a region of higher density towards the bottom of the plot, representing peroxisomes that are predominately CFP. The normalization protocol took the CFP and YFP data and skewed the ratio towards CFP due to the normalization calculation lacking an iteration step. This is entirely against the reason for normalizing the data. It was assumed that changing the scale of the data proportionally by performing the same mathematical operations would preserve the YFP/CFP scaling. This is not the case. For this reason, the dual channel data was not considered during the analysis.

6.7 Matlab Analysis of Tri-channel CFP-YFP-RFP Data

Once the analysis of the dual channel data was determined to be insufficient, the script was adapted to analyze the images for quantitative RFP information along with CFP and YFP. This script is in Appendix 8: Matlab image processing script. The script was modified to perform the same analysis on RFP that was already performed on CFP and YFP. The addition of a common

reference frame in the form of the RFP data was used to normalize the CFP and YFP without changing the scaling information in favor of one channel or the other. The normalization was first performed on an image to image basis and finally across the data-set as a whole.

6.7.1 Example Analysis Strain: KJY15

The script used in this section contains additional code to recorded quantitative data on the RFP peroxisomes. These data were used as a common reference for the CFP and YFP. The individual values of the CFP and YFP peroxisomes would only be changed relative to the scale of the RFP. This process allowed the data to be examined without changing the scale of the individual peroxisomes.

The first sample to be analyzed is the KJY15 strain that has been processed using macro V5. As previously described, KJY15 was an ideal initial sample to examine as it has both CFP and YFP fluorophores where both are being driven by the yeast *PEX3* promoter. These data represent as close to a normal distribution as has been observed with the control samples. Figure 6.12 through Figure 6.17 illustrates the normalization protocol that has been written in Matlab.

6.7.2 Image to Image Normalization

This section details the normalization procedure in a similar way to the method presented for the CFP and YFP information. A series of figures are presented where the left plot contains a smoothed image to image average and the right plot shows the changing image average intensity.

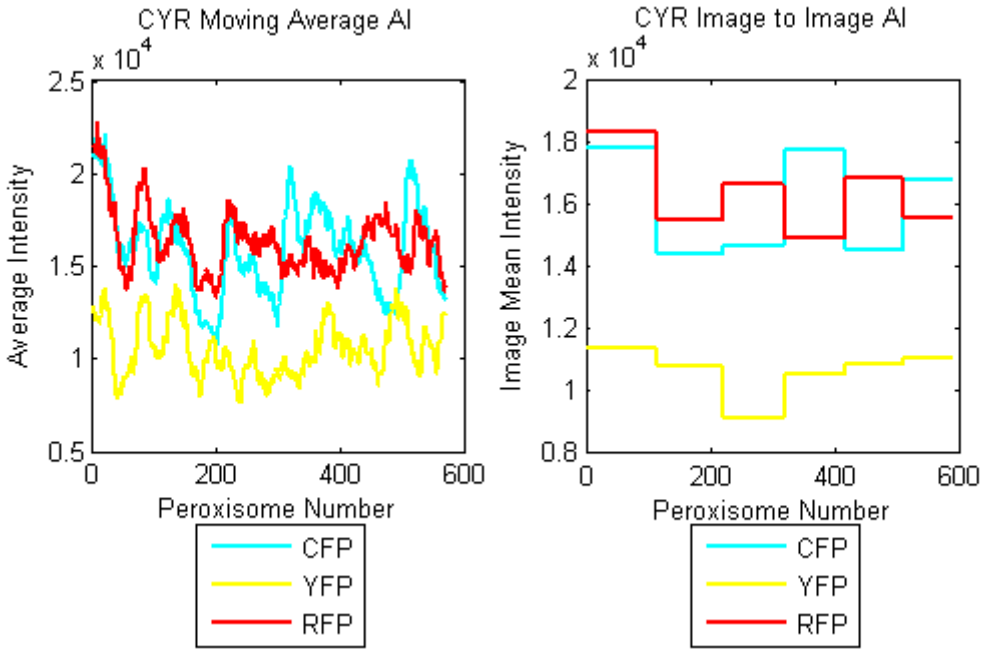


Figure 6.12: CYR (Cerulean, mVenus and mCherry) raw data moving average and image to image average intensity. This data was acquired from 6 images of the KJY15 strain (*pex3-Δ, pex19- Δ*) then processed through scripts in Image Pro7 and Matlab.

The first step in the normalization protocol was to examine a moving average of the CYR average intensity data. This was shown in the first plot in Figure 6.12. The second plot in the same figure shows the image to image average across the experiment. In this case, there are 6 plateaus, indicating that six image have been used in this analysis.

This figure shows the RFP and CFP channels are on approximately the same scale but the YFP channel is much lower in intensity. This phenomenon is dependent on the fluorescent proteins and is simply part of imaging biological samples. Using a 16 bit camera with and EMCCD detector will also have an effect on the scale of the data. The acquisition protocol was setup to capture information based on the maximum intensity peroxisome. YFP is a stronger fluorophore than CFP but the YFP average is lower as the single greatest YFP value is much higher than the single greatest CFP value. The result of this is effect is a higher CFP average despite CFP being a less sensitive fluorescent protein.

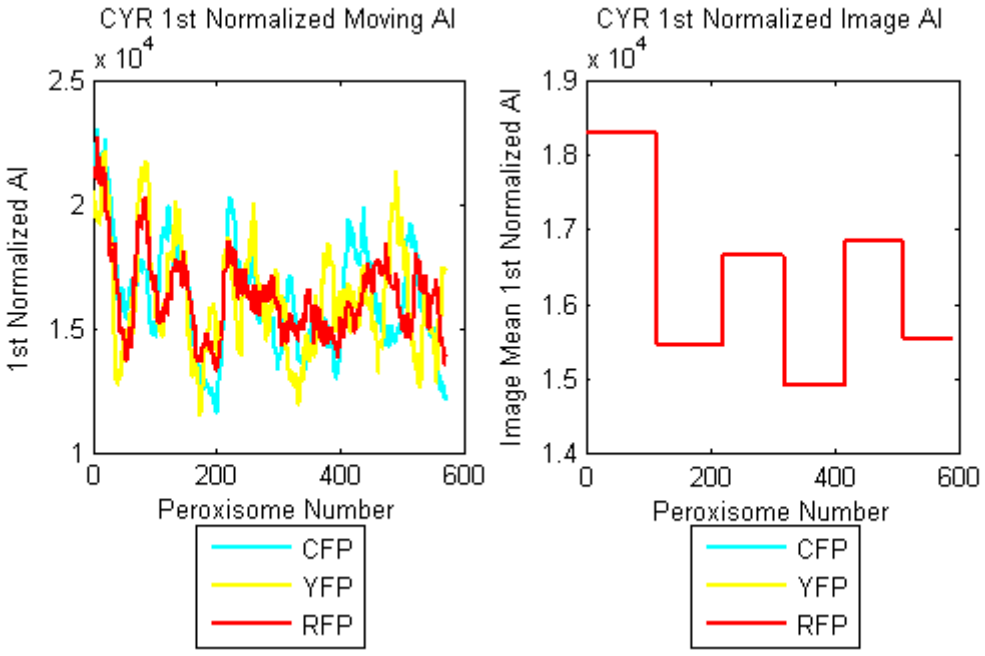


Figure 6.13: CYR (Cerulean, mVenus and mCherry) first normalized moving average and image to image first normalized average intensity. The image average intensity (Image AI) plot shows only one line, as the CFP and YFP image to image averages have been normalized to the RFP average.

The first part of the normalization protocol adjusts the CFP and YFP image averages to the RFP average. The first graph in Figure 6.13 shows the image to image moving average after this normalization has been performed. The second plot in the figure illustrates how the image to image average now appears. Note that the only line that appears is RFP as it is the final entry in the legend; it covers over the other two colors.

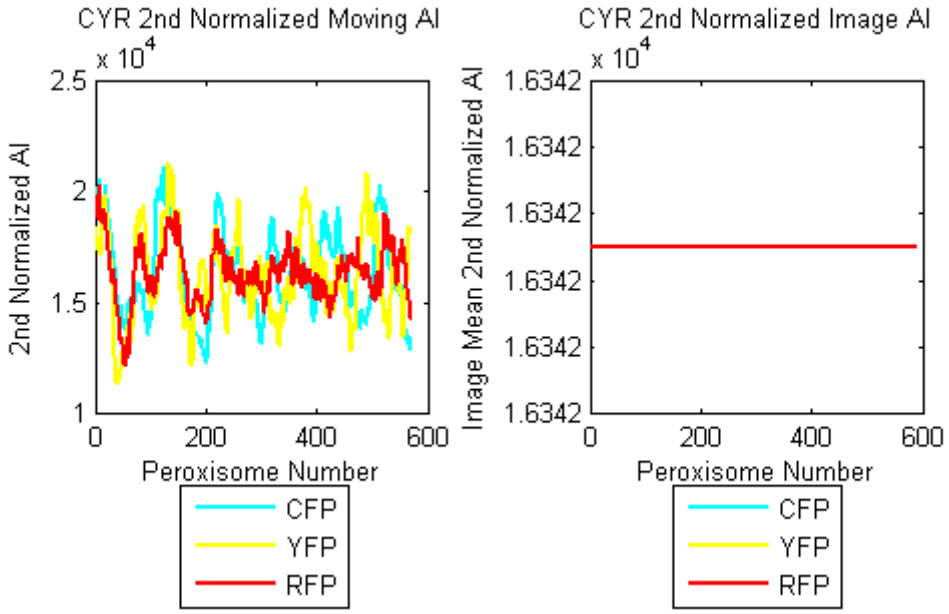


Figure 6.14: CYR (Cerulean, mVenus and mCherry) second normalized moving average and image to image second normalized average intensity. The second normalized average AI shows only one line, as the image to image average for the image-set has been normalized to the average of the RFP channel.

Figure 6.14 illustrates the final step in the normalization protocol. The image to image average is on the same scale across all images, as shown by the second plot in the figure. The image to image average is now constant and the scale of the CFP and YFP values can now be compared. This normalization protocol has removed any channel specific bias that might have occurred during the imaging without changing the resolution of the data.

6.7.3 Scatter-plot and Histogram Analysis

The second part of the normalization script will show the scatter plot and histogram. Three figures will be shown that highlight the changes in the shape of the histogram and the scatter plot with each stage of the normalization protocol.

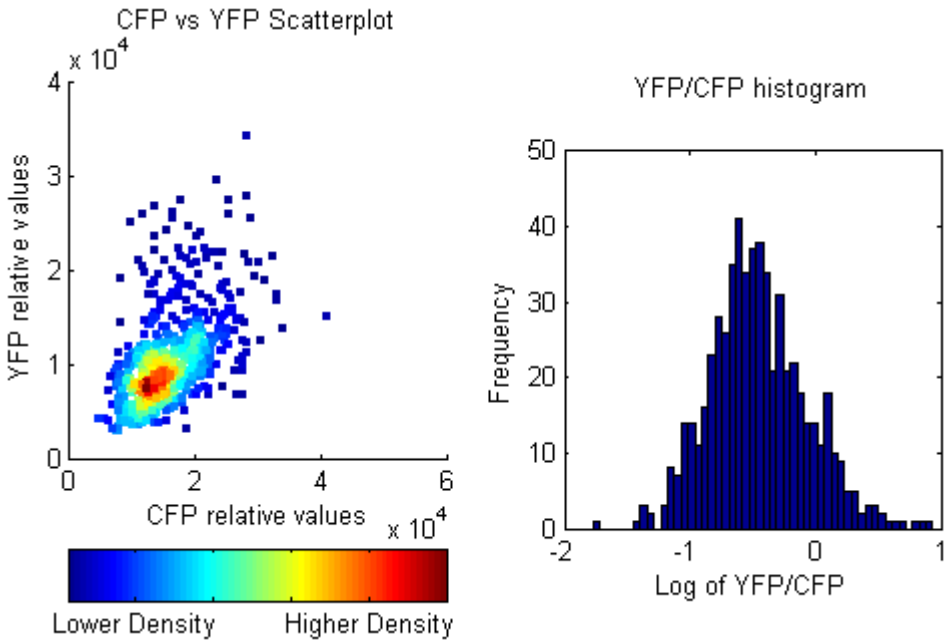


Figure 6.15: CFP vs. YFP average intensity scatter plot and YFP/CFP histogram. The scatter plot clearly shows a single area of higher intensity and the histogram also shows a single distribution.

The first graph in Figure 6.15 shows a scatter plot of the raw YFP and CFP average intensity data. The color bar shows the relative density of the points. The second plot in the figure is a histogram of the log of the YFP/CFP peroxisome average intensity raw data. These two plots represent two different ways of looking at the same data. The scatter plot shows how the two variables are related and the histogram shows the distribution of YFP/CFP ratio.

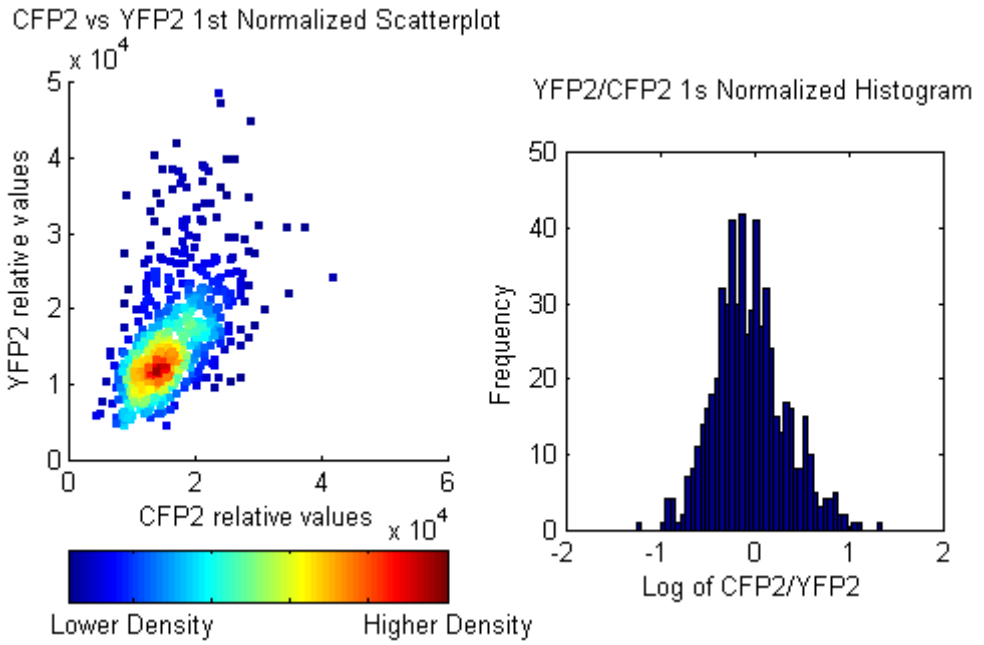


Figure 6.16: CFP vs. YFP first normalized average intensity scatter plot and YFP/CFP first normalized histogram.

Figure 6.16 shows the CFP vs. YFP scatter plot and the YFP/CFP histogram of the first normalized data. The shape of the scatter plot and histogram may change a little in this intermediary normalization, but it is the final normalization that will be analyzed.

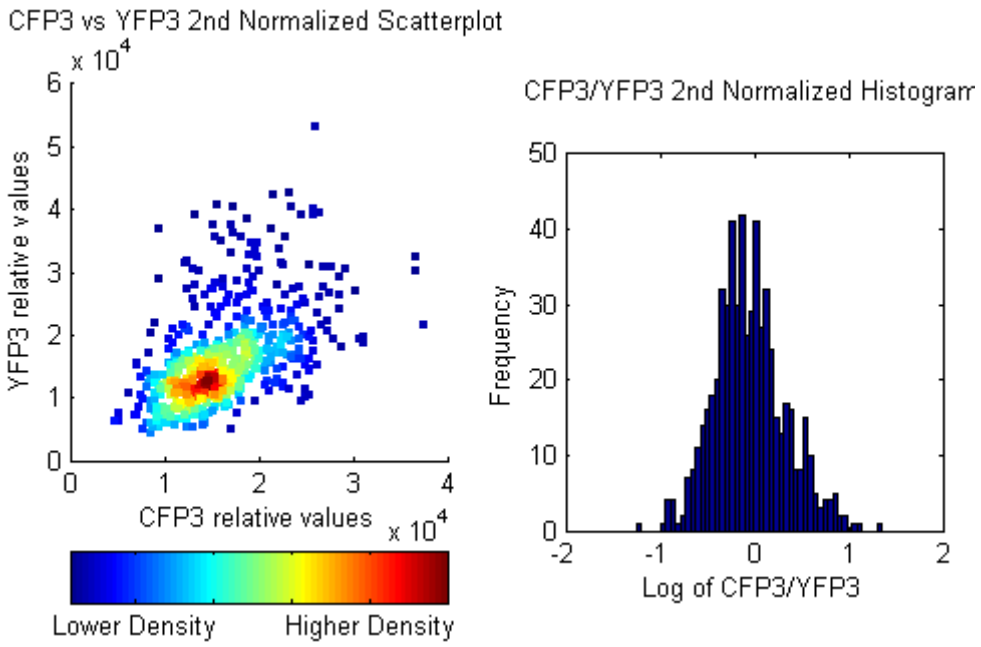


Figure 6.17: CFP vs. YFP second normalized average intensity scatter plot and YFP/CFP second normalized histogram. The shape of the scatter plot has changed relative to the raw data, but the plot still shows a single area of higher density.

Figure 6.17 shows the second normalization of the CFP3 vs. YFP3 scatter plot and the log of the YFP/CFP histogram. These data were analyzed to probe the relationship between YFP and CFP peroxisomes when normalized to the same scale. Note the histogram in Figure 6.16 is identical to Figure 6.17 as the scaling element from the image to image normalization is already corrected with the YFP/CFP division. There is a change in the scatter plot between Figure 6.16 and Figure 6.17 because of the image population average normalization; hence the latter will be used for analysis.

The script was also initially set to plot the R^2 value for the three scatter plots. After some initial investigation, it was clear that looking at R^2 values for a scatter plot was insufficient as fitting a line to data that is distributed over an area was leading to analytical artifacts. The scatter plot was chosen to be examined for multiple populations using a heat map, while the histogram was used for curve fitting analysis.

6.8 Histogram Curve Fitting Analysis

The histograms were analyzed using curve fitting algorithms in the Matlab statistical analysis toolbox. Three distributions have been chosen:

1) Normal distribution
2) T-location scale
3) Non-parametric

These distributions were chosen to represent a good cross section of possible fits for the data. The normal distribution was chosen as the KJY15 strain would be expected to fall closest to this distribution. The T-location scale distribution was used to model data with heavier tails or more outliers than a normal distribution. Finally, the non-parametric distribution with a 0.25 weighting was chosen to fit bimodal distributions. This distribution calculates a probability density estimate of the sample based on weighted moving average. If the weight is set to 1, the algorithm tries to fit the data to a normal distribution. If this weight is changed a smoothed average can be plotted. A weight of 0.25 was used as it gives enough resolution to show heavy tails and multiple modes in the data.

6.8.1 Histogram Distribution Visualisation Fit Analysis

Each distribution was fit to the CFP-YFP normalized histogram. Matlab uses an iterative curve fitting algorithm to fit the chosen distribution as closely as possible to the data. Once this fit was calculated, the plot was mapped as shown in Figure 6.18.

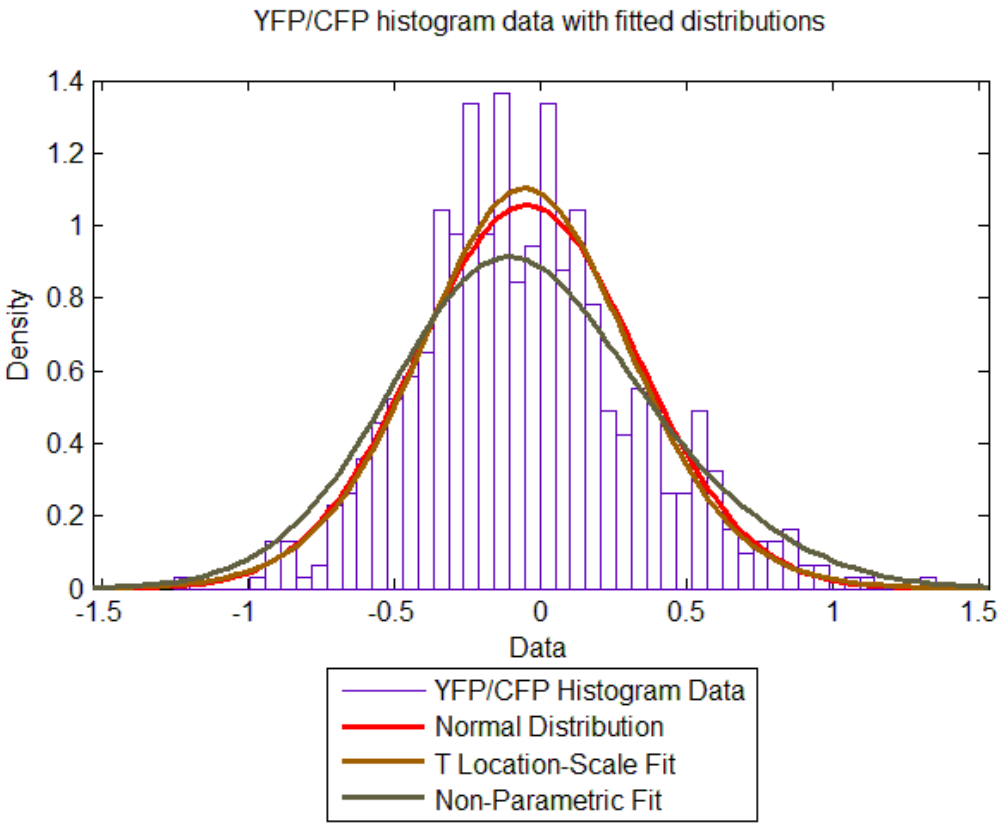
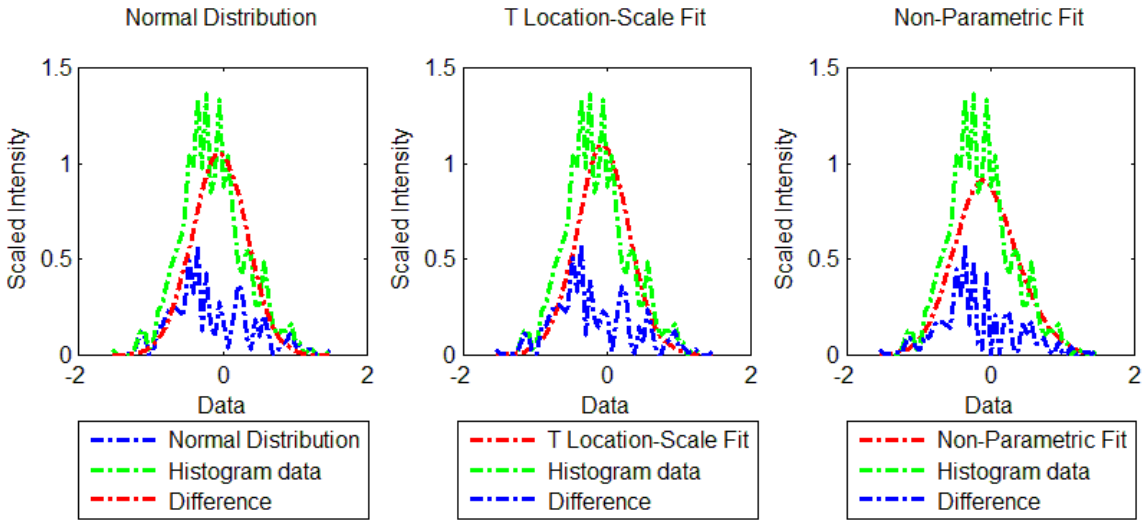


Figure 6.18: KJY15 YFP/CFP histogram with fitted statistical distributions as calculated by Matlab.

An initial visual inspection indicates that although the non-parametric fit appears the least compatible with the data, it was impossible to discern the best fit between the other distributions. A numerical analysis of the fit was conducted to assess which distribution most closely matches the data.

6.8.2 Distribution Fit Analysis

To assess the curve fit, the distribution was subtracted from the data and the sum of the squared residuals calculated. In this case, a lower value indicated a better fit. A value of 0 indicated that Matlab has perfectly matched the curve to the data although this outcome was not expected.



A) **B)** **C)**

Figure 6.19: Assessment of the normal, T location-scale and non-parametric fit when compared to the YFP/CFP histogram data. The histogram has been re-plotted as a line graph for clarity, as represented by the green line. The red line represents the distribution and the blue, the difference between the two. A numerical assessment of distribution fit is assessed by calculating the square of the numerical value of the blue line.

Figure 6.19 shows A the normal distribution, B the T-Location scale and C the Non-parametric curve fitting plots where the red line is the distribution, the green line is the histogram data plotted as a line graph and the blue line is the difference. The histogram was replotted as a line graph to illustrate the residual analysis. The blue line represents the difference between the fit and the histogram data and the square of the values for each point is calculated to assess the fit numerically. The distributions superimposed on the histogram data were shown in Figure 6.18. As explained previously, the T-location scale distribution was used to model data with heavier tails or more outliers than a normal distribution and the non-parametric distribution was used to identify a bimodal distribution.

KJY15 strain distribution assessment	
Normal Distribution	1.77
T Location-Scale Fit	1.78
Non-Parametric Fit	1.48

Table 6.4: KJY15 strain distribution fitting numerical results

Table 6.4 shows the sum of the residuals squared for the KJY15 strain. The lowest value is the non-parametric fit. It also indicates that the best fit for the YFP/CFP histogram was that of a bimodal distribution.

6.8.3 Matlab Analysis Summary Table

Table 6.4 shows a summary of the Matlab operations used to conduct the numerical analysis

.

Matlab analysis
Load in Excel data
Parse CFP, YFP and RFP channels
Normalize CFP and YFP image to image intensity to RFP channel
Normalize CFP and YFP images average intensity to population average
Plot CFP vs. YFP scatter heat map of normalized average populations
Plot log of CFP/YFP histogram of normalized data
Fit distributions to histogram
Assess distribution fit numerically and produce quantitative value of fit

Table 6.5: Matlab comparison summary table

6.9 BY4741a Scatter Plot and Histogram Analysis

This numerical technique was written to compare two populations of peroxisomes that were normalized based on a third channel. Using a third channel served two functions. The first was to normalize that data on an external element different to both channels and the second was to normalize based on functional peroxisomes. If the peroxisomes were deficient in the uptake of the tagged RFP molecule, they did not show up on the RFP image and hence were excluded from the analysis. This technique only worked with samples that contain CFP, YFP and RFP. Only KJY11, KJY14, KJY15 and the KJY10 (main experimental strain) meet this criteria and were be analyzed for quantitative data. The other control strains in the project served their purpose by demonstrating that the constructs worked and could rescue peroxisome deletions or by producing proof of concept images.

6.9.1 Analysis: KJY11

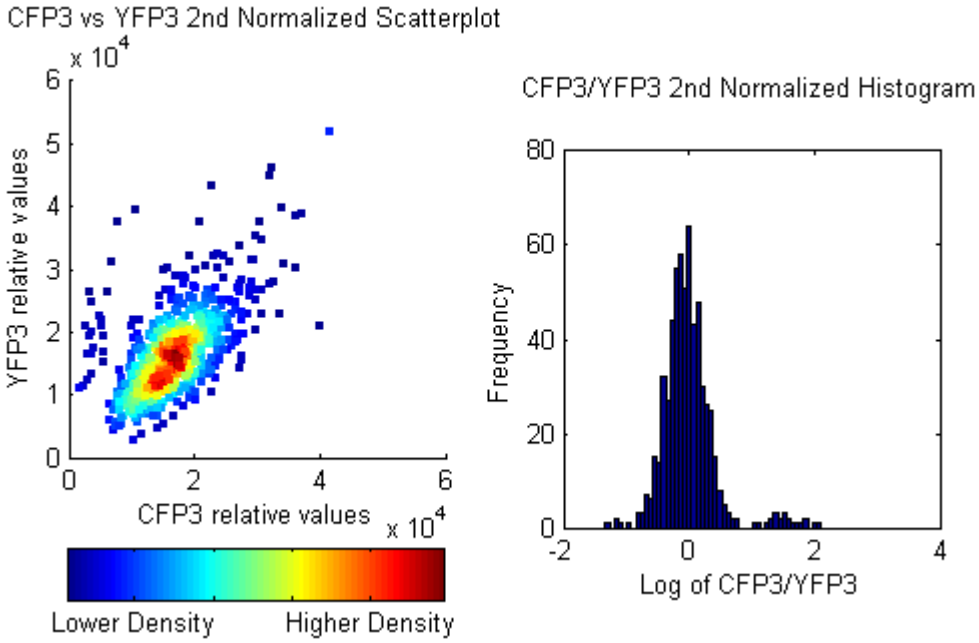


Figure 6.20: Normalized YFP/CFP scatter plot and histogram for the KJY11 control strain. The scatter plot shows a few isolated points closer to the YFP axis, and the histogram shows tails offset from the main population that represent these points.

The KJY11 control strain is missing yeast *PEX19* and would therefore be expected to show a bias towards YFP-human peroxisomes. The scatter plot heat distribution does not indicate the presence of two populations of peroxisomes. The histogram shows one large peak with a small asymmetric tail.

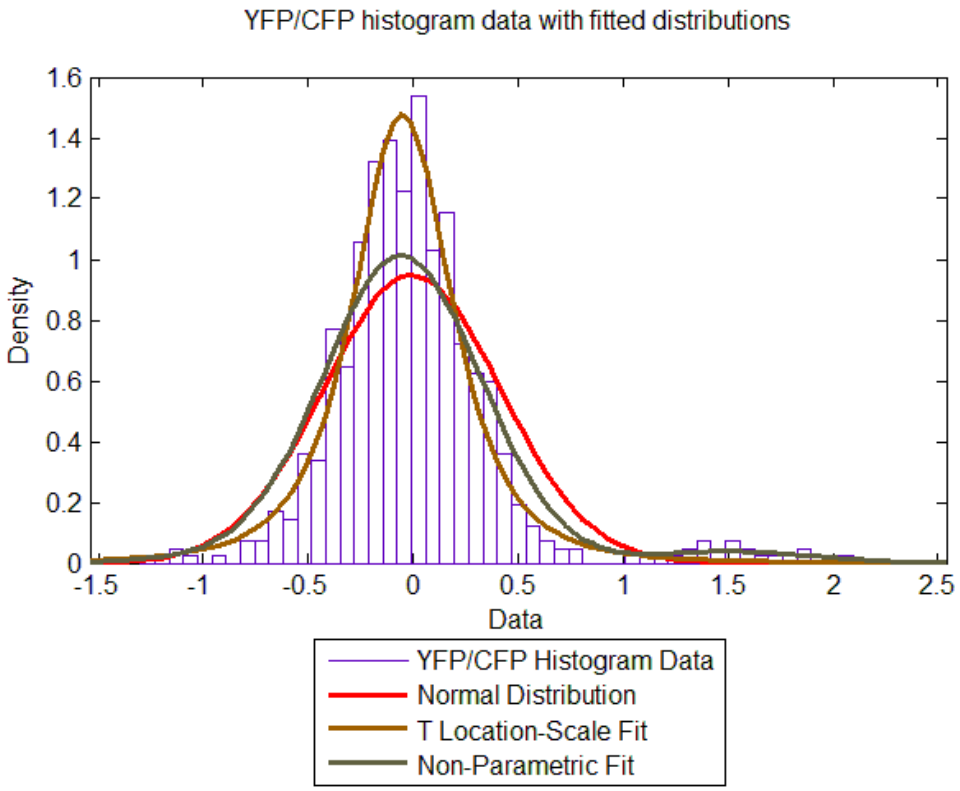


Figure 6.21: KJY11 YFP/CFP histogram with fitted distributions.

KJY11 strain distribution assessment	
Normal Distribution	1.28
T Location-Scale Fit	0.71
Non-Parametric Fit	0.97

Table 6.6: KJY11 strain distribution assessment.

Figure 6.21 shows the distributions that have been fitted to the histogram and Table 6.6 shows the distribution summary. Visually, the best distribution is the T location-scale fit, which is confirmed with the numerical fit assessment. This distribution fits best to asymmetric data with long tails.

6.9.2 Analysis: KJY14

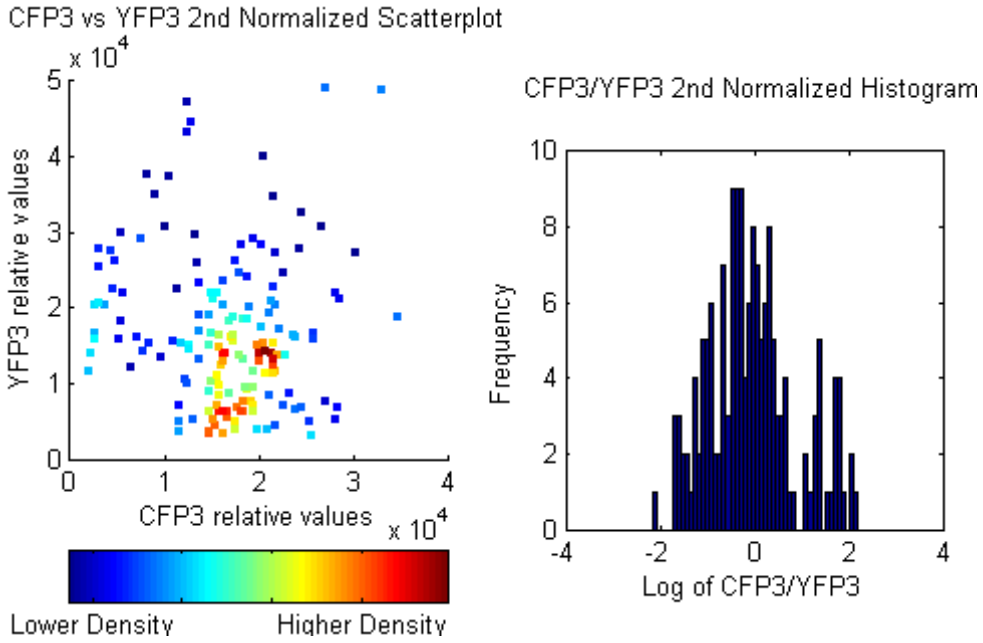


Figure 6.22: Normalized YFP/CFP scatter plot and histogram for the KJY14 control strain.

Figure 6.22 shows the YFP/CFP scatter plot and histogram for the KJY14 control strain. The scatter plot immediately shows little coherence or pattern to the peroxisome distribution. It also indicates fewer peroxisomes than in the previous plot and the histogram is of lower resolution. This strain has CFP and YFP tagged to yeast Pex3p with human and yeast Pex19p. This strain was expected to exhibit a histogram closely resembling a normal distribution. Although the points are sparse, the scatter plot shows two different regions of higher density.

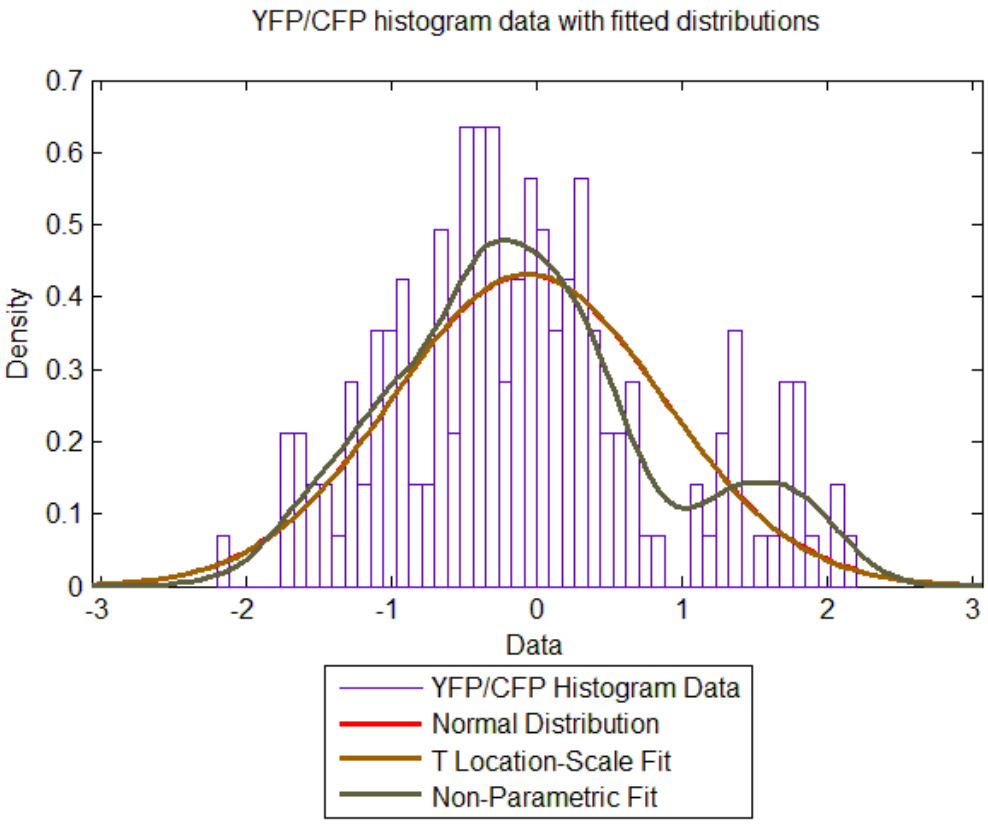


Figure 6.23: KJY14 YFP/CFP histogram with fitted distributions. Note the normal distribution was calculated but not plotted due to Matlab reaching the maximum number of iterations without producing a curve that could match the data.

KJY14 strain distribution assessment	
Normal Distribution	1.08
T Location-Scale Fit	1.08
Non-Parametric Fit	0.98

Table 6.7: KJY14 strain distribution assessment.

Figure 6.23 shows the distributions fitted to the KJY14 control strain and Table 6.7, the numerical fit summary. The first observation is that the normal distribution is missing from the plot. This is due to the algorithm in Matlab being unable to fit a normal to the data within the standard number of iterations. As the curve for this distribution was unable to be fitted to the plot, it will not be considered for analysis (although a numerical assessment of the fit will still be calculated). Visually, the best distribution was the non-parametric fit as it maps the two major

modes. This was confirmed with the non-parametric fit being the lowest numerical result. However, the images for this strain contained a large degree of cytoplasmic localization, rendering numerical analysis difficult at best.

6.9.3 Analysis: KJY15

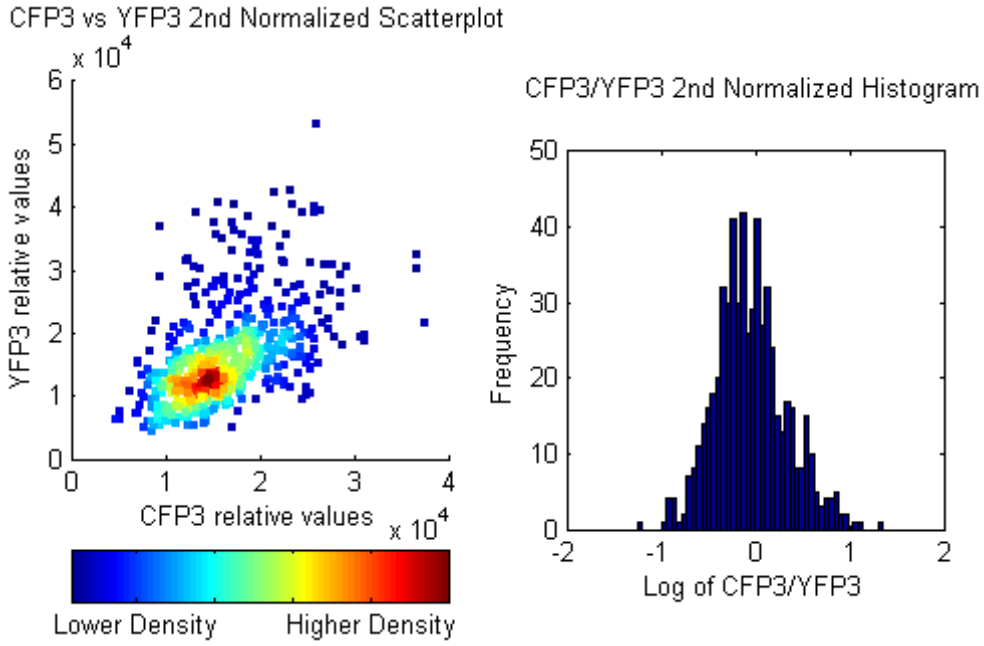


Figure 6.24: Normalized YFP/CFP scatter plot and histogram for the KJY15 control strain.

Figure 6.24 shows the scatter plot and histogram for the KJY15 control. This control has CFP and YFP both tagged to the yeast Pex3p and two copies of the yeast Pex19p.

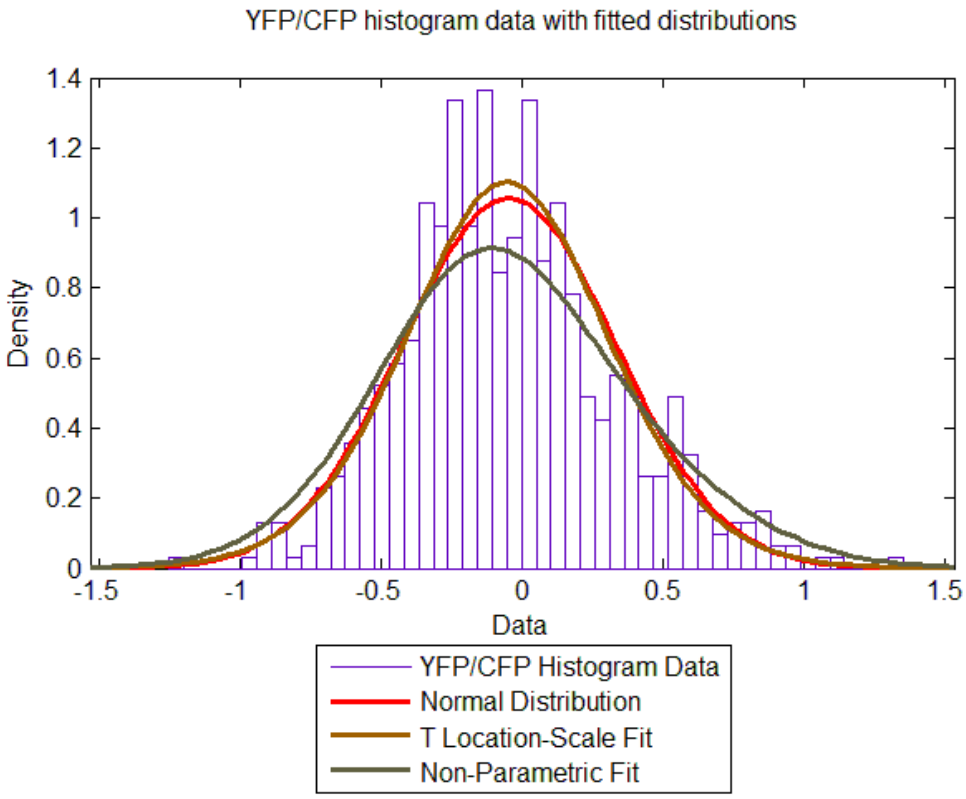


Figure 6.25: KJY15 YFP/CFP histogram with fitted distributions.

KJY15 strain distribution assessment	
Normal Distribution	1.77
T Location-Scale Fit	1.78
Non-Parametric Fit	1.48

Table 6.8: KJY15 strain distribution assessment.

Figure 6.25 shows the fitted distributions for the KJY15 control and Table 6.8 gives a summary of the numerical fit assessment. The non-parametric distribution was numerically calculated to be the best fit. This strain was expected to exhibit the characteristics of a T-location scale distribution.

6.9.4 Analysis: Main Experimental Strain KJY10

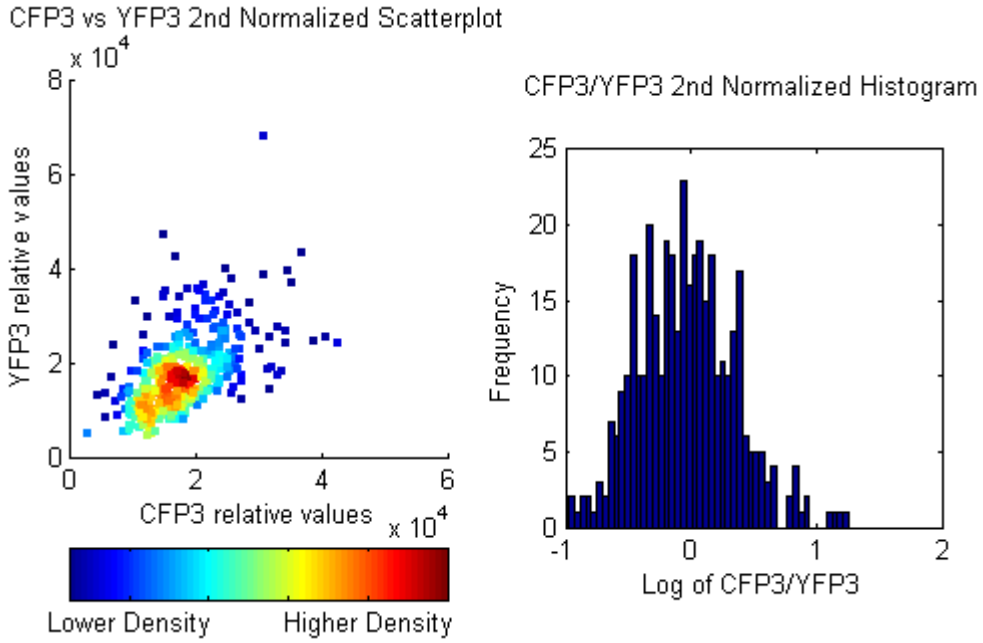


Figure 6.26: Normalized YFP/CFP scatter plot and histogram for the KJY10/main experimental strain.

Figure 6.26 shows the second normalized YFP/CFP scatter plot and histogram. There is clearly only one area of points at high concentration as shown by the red color. The histogram only shows one large main peak, although this will be analyzed by the curve fitting.

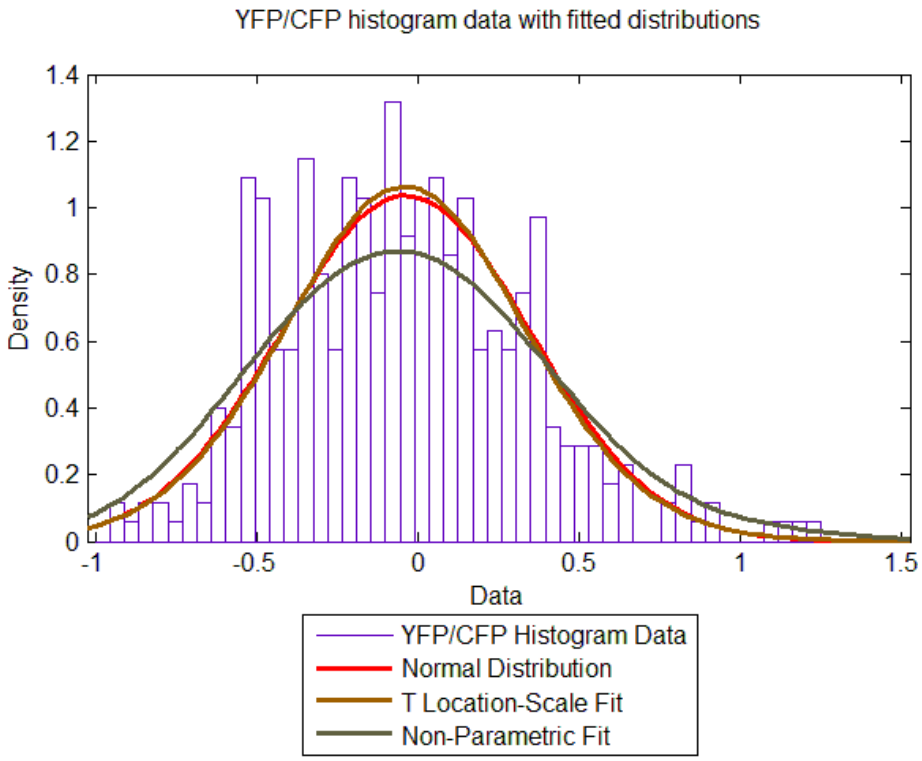


Figure 6.27: KJY10 YFP/CFP histogram with fitted distributions.

KJY10 strain distribution assessment	
Normal Distribution	1.48
T Location-Scale Fit	1.55
Non-Parametric Fit	1.61

Table 6.9: KJY10 strain distribution.

Figure 6.27 shows the YFP/CFP histogram data and fitted distributions and Table 6.9 gives the numerical fit summary. The KJY10 strain contains the human Pex3p tagged with CFP and the yeast Pex3p tagged with YFP. One of two results was expected from this experiment. This strain was expected to either show two separate populations of human and yeast peroxisomes in the cell, or one homogenous population of hybrid peroxisomes. The scatter plot shows only one population and the best distribution to fit the data is the normal distribution. These data suggest the experiment produced a single population of human-yeast hybrid peroxisomes.

6.9.5 Numerical Summary Table

The previous section presented the histogram distribution and scatter plot analysis. This section compares the numerical results of the four distributions for the three control strains and the experimental strain.

Matlab Data Summary					
Strain	KJY15	KJY11	KJY14	KJY15	KJY10
Date imaged:	02-Dec-09	01-Dec-09	01-Dec-09	01-Dec-09	01-Dec-09
The number of peroxisomes in all images =	497	613	161	591	388
The number of cells in all images =	325	621	312	510	270
Peroxisome/Cell ratio =	1.53	0.99	0.52	1.16	1.44
Number of images in this data set =	4	7	6	6	7
Sum of the squares of the residuals from curve fitting:					
Normal Distribution	2.26	1.28	1.08	1.77	1.48
T Location-Scale Fit	2.25	0.71	1.08	1.78	1.55
Non-Parametric Fit	2.80	0.97	0.98	1.48	1.61

Table 6.10: Matlab data summary.

Table 6.10 shows a summary of the distribution analysis. The best fit distributions are highlighted in yellow. As was explained earlier in the Chapter, the yellow boxes represent the closest fit. KJY15 was expected to be a normal distribution as this strain had two copies of the yeast Pex3p, one tagged with CFP and the other with YFP. The other point of interest in this table is the KJY10 strain data fitting a normal distribution.

6.10 Chapter Conclusions

This Chapter presented image and numerical analysis of the peroxisomes using scripts written in Matlab to produce scatter plots and YFP/CFP histograms. The dual channel data was analyzed and a normalization script was written to normalize the data from image to image, then across the data set. This protocol required an iterative numerical solution to complete the second normalization which was deemed too complex for this analysis. The scatter plot and histograms were examined and the technique deemed insufficient due to the change in the YFP to CFP scaling ratio.

The dual channel data using YPH500α was then examined and a flaw in the original technique discovered. This flaw was theorized to come from an observer introduced bias in the data capture. The dual channel images were therefore removed from the analysis.

The tri-channel data was analyzed using a similar normalization protocol that made use of the additional RFP image. This script normalized the CFP and YFP channels to the RFP channel, then normalized the images across the data-set. The second normalized histogram was then analyzed using the curve fitting algorithm in Matlab. Three distributions were chosen to represent the possible situations: normal, T-location and non-parametric.

The three key control strains and the main experimental strain were analyzed and the second normalized log of the YFP/CFP histogram fitted to these distributions. The KJY14 strain fit the non-parametric distribution the best. This distribution is used to fit bimodal distributions. The main experimental strain KJY10, was matched the closest to a normal distribution. This strain was expected to either have a bimodal distribution if two populations of peroxisomes were present or a normal distribution if this population was homogenous. This was corroborated by the scatter plot showing no evidence of two populations. It is therefore possible to conclude that the human and yeast Pex3p work closely in tandem, producing a uniform population of yeast-human hybrid peroxisomes.

7 Discussion and Further Work

The aims of this Thesis were to develop a standard to measure the activity of *E. coli* promoters and compare results to those of other labs using an *in vivo* reference standard, to develop and build a set yeast strains capable of expressing human and yeast peroxisomes, to image these yeast strains, to validate the nature of the peroxisome population through numerical analysis, and finally, to investigate the creation of heterologous (or artificial) peroxisomes.

An *in vivo* reference standard to measure *E. coli* promoter activity was created using BioBricks from the Registry of Standard Biological Parts. This standard was called the Relative Promoter Unit (RPU) and was based on comparing the activity of a reference promoter BBa_J23101 to other promoters under the same conditions. This standard was used to measure the activity of a collection of promoters that were distributed to multiple labs at six different institutions. The results showed a less than two-fold range across the three promoters measured by the labs, when standardized to the reference promoter. This activity shows that a lack of standardized equipment and calibrations does not inhibit recording standard measures of promoter activity when using a reference standard and protocol. Labs can now standardize promoter measurement by simply adhering to a measurement protocol and using a reference strain.

The aim of the yeast promoter experiments was to assess the promoter activity of a strain collection in several different media compositions by applying knowledge gained from the *E. coli* measurement experiments. Further work needs to be conducted to measure the promoter collection strains with single constructs, in a standard glucose media. Using the Kelly method with only the YFP fluorophores could result in a more robust, reliable comparative measure for yeast promoters where the *ADH1* promoter could be used as the reference standard.

A collection of yeast strains was built in an attempt to create artificial peroxisomes. Background strains (BY4741a and YPH500α) were selected and the key peroxisome biogenesis proteins (Pex3p and Pex19p) were knocked out. BioBrick constructs were assembled from yeast and

human genomic DNA to rescue the function of these peroxisomes. These constructs were built and assembled in *E. coli* and then transformed into yeast to create a collection of control strains and a master experimental strain. The BY4741 strain eventually proved more successful and was used to collect data. Once all the controls and the main experimental strain had been constructed, the imaging and analysis was conducted.

This Chapter highlights the key findings from the results and analysis Chapters. The *E. coli* and *S. cerevisiae* promoter measurement work is discussed, followed by the peroxisome project. Key microscope setup information, images and numerical results are interpreted and discussed in the greater context of the work. Chapter 5 presented images of the peroxisome strains. Chapter 6 showed the results of the numerical Matlab analysis. Finally, this work will be discussed in the greater context of the scientific community.

7.1 E. coli *and* S. cerevisiae *Promoter Measurement Discussion*

A method to measure the promoter activity in *E. coli* across multiple labs using a standardized reference promoter was presented. Construct BBa_I20260 (promoter only part: BBa_J23101) was chosen as the constitutive *E. coli* reference promoter. Results from the multi-institution experiments were shown, indicating that using a relative standard and measurement protocol can produce close results across labs with non-standard equipment. This method built on previous work by Miller by allowing promoter measurement to be compared between labs on different days (Miller and Zipser 1977). It was hoped researchers will adopt this method to increase the body of knowledge associated with the characterization of biological parts. To this end, a measurement kit was developed and distributed to the iGEM teams during the 2008 competition and discussions with teams who used the kit reported that it worked well.

The aims of the yeast promoter experiments were to assess the promoter activity of a strain collection in different media compositions by applying knowledge gained from the *E. coli* measurement experiments and to investigate a yeast reference promoter. This work indicated that the *ADH1* promoter could be used as a reference promoter. Selecting this promoter must, however, come with the caveat of acquiring measurements during mid-log phase when the strain is grown in glucose. The promoter is known to be repressed as the concentration of ethanol in the medium increases which will change the nature of the measurements. The protocol specified

glucose as the carbon source in the media. This promoter should be tested and characterized relative to other suitable candidates such as *HXT1*, *HXT2*, *HIS3* and potentially *PGK1* in order to select the correct standard promoter in yeast. *HXT1* and *HXT2* regulate hexose transport, allowing yeast to grow on glucose (Reifenberger, Freidel et al. 1995). They are responsive to different concentrations of glucose, therefore, further testing would be required to assess which of the two promoters would be better suited to measurements over time. *HIS3* is involved in histidine biosynthesis, so using it as a measurement standard would require standard growth conditions to regulate the quantity of histidine in the media. *PGK1* is involved in ATP biosynthesis. It is highly expressed, the mRNA is highly stable and has been thoroughly studied by the yeast community (Chambers, Tsang et al. 1989; Packham, Graham et al. 1996). The ideal solution would be a synthetic promoter that expressed the same protein irrespective of growth media, conditions or growth phase and potentially organism. However, designing this promoter would present a significant technical challenge, irrespective of whether this feat is actually possible.

The first step in assessing these promoters would be to test how their expression levels change over time during different phases of yeast growth, which reflect dynamic changes in nutrient availability as culture medium is progressively depleted. If they express protein at a constant rate during mid-log phase, then protocols from the Kelly method could be employed. If not, they may not be suitable to use as a standardized promoter for yeast. Further work is required to measure the promoter collection strains with single constructs. YFP measurements were ultimately not reported in conditions other than SD. Using the Kelly method with only the YFP fluorophores could result in a more robust, reliable comparative measure for yeast promoters.

7.2 Synthetic Hybrid Peroxisome Discussion

The KJY14 control strain produced images with cytoplasmic localization in the YFP channel, a result that was unexpected (genotype: BY4741 *MAT*a, *pex3-Δ, pex19-Δ,* LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-hsPEX19, HIS1::pPEX3-scPEX3-YFP-ADH1t-pPEX19-scPEX19). The yeast Pex3p-YFP construct was locating to the cytoplasm in some cases (approx. less than 20% of cells). This strain possessed both human and yeast copies of Pex19p, so interference could be hypothesized to come from the human Pex19p. If the human Pex19p docked to yeast Pex3p and did not undock the cargo, there could be an

accumulation of Pex3p proteins docked to human Pex19p in the cell, which might appear as cytoplasmic localization. This problem could be assessed and monitored if time-lapse videos, cell tracking and peroxisome tracking had been implemented.

The KJY10 strain (BY4741 MATa, pex3-Δ, pex19-Δ, LEU2::pHIS3-RFP-RFP-PTS1, TRP1::pPEX3-scPEX3-CFP-ADH1t-pPEX19-scPEX19, HIS1::pPEX3-hsPEX3-YFP-ADH1t-pPEX19-hsPEX19) was the main experimental strain for the work conducted on the DeltaVision. It was transformed with three constructs. The first was the RFP-PTS1 to test for peroxisome function. The second is the yeast Pex3p-CFP and Pex19p construct, to induce yeast-CFP peroxisomes. The third is the human Pex3p-YFP and Pex19p construct to induce human peroxisomes. Figure 5.28 A) shows the RFP-PTS1 construct and functional peroxisomes. Figure 5.28 B) shows the yeast Pex3p-CFP peroxisomes. Figure 5.28 C) shows the human Pex3p-YFP peroxisomes. Figure 5.28 D) is the first false color composite CFP-YFP image. As explained in the introduction to this Chapter, CFP is represented by green and YFP by red with yellow being the overlap of the two colors. This image shows most of the peroxisomes appearing as yellow/light green indicating that there is little evidence of two species specific populations of peroxisomes. Figure 5.28 D shows a very close correlation to both Figure 5.28 B (Cerulean) and Figure 5.28 C (mVenus) upon visual inspection. This information has been obtained using a much better suited microscopy setup, an observer bias free image analysis technique and a superior 3d-deconvolution process. Images from the DeltaVision are better representations of the actual depth of grey scale.

Automated image analysis of peroxisomes has been previously described (Niemisto, Selinummi et al. 2006). This analysis was furthered by comparing two populations of peroxisomes based on a common marker for peroxisome function (RFP-PTS1) in search of bimodality. Figure 6.26 shows the second normalized YFP/CFP scatter plot and the log of CFP/YFP histogram. On the scatter plot, there is clearly only one area of high density points as shown by the red color. This denotes a single population of peroxisomes with uniform characteristics. Figure 6.27 shows the YFP/CFP histogram data and fitted distributions. The normal distribution was found to have the best fit by calculating the sum of the residuals squared. This strain was expected to either have a bimodal distribution if two populations of peroxisomes were present or a normal distribution if

this population was homogenous. A normal distribution was observed indicating a lack of bimodality which was corroborated by the scatter plot also showing no evidence of two populations of peroxisomes.

The script to analyze the image data collected using the Nikon microscope was flawed due to an error in the observer thresholding bias. This error was introduced due to the lack of third channel (RFP) to allow for normalization. The error introduced a significant bias on the bimodality due to the image to image variation. When spurious images were removed from the data-set, the bimodality disappeared, leading to the knowledge that this analysis technique was insufficient. This applied to all dual channel data that was processed with this technique. The DeltaVision analysis technique starts with an image to image normalization process as has been explained in Chapter 6. This process results in a change to the relative scaling of the quantified data, leaving the only comparison metric to be the shape and nature of the distribution on the histogram and the pattern of the scatter plot. The relative scale of the data cannot be compared as it has been normalized on an image to image basis. As the important quantities are the distribution of the CFP/YFP histogram and the shape of the scatter plot, this was not deemed a problem.

When investigating such a phenomenon in the future, several important questions first need to be asked. The first is whether the microscopy setup is correct and sufficiently sensitive for the expected phenomenon? Does the camera capture all the observed phenomena, i.e. can the sensor be saturated without using unreasonably long exposure times and extreme levels of brightness? When examining the analysis, what are the most important points relative to the data needed to demonstrate the phenomena under investigation? What factors will make this analysis more or less robust and what factors is it most sensitive to? Contemplating and answering these questions in future projects would likely make data capture and analysis more robust and reliable at the stage in the project where changes are easy to make.

7.3 Conclusions

Previous research has been conducted to investigate the construction of artificial organelles. One such group sought to recreate the function of the Golgi apparatus *in vitro* using microfluidics, magnetic particles and recombinant enzymes (Martin, Gupta et al. 2009). Another

group created artificial organelles *in vivo* using polymers that were internalized and remained stable in macrophage cells for up to 48 hours (Ben-Haim, Broz et al. 2008). The 2007 UCSF iGEM team set out to make an artificial organelle by expressing a higher eukaryote lipid phosphatase MTM in *S. cerevisiae*. They were successful in their attempts and produced fluorescence images of GFP tagged synthetic organelles in yeast (Chen 2007). However, none of the groups mentioned have looked at using peroxisomes as a base for organelle engineering. We built a yeast strain that produced novel functional peroxisomes in *S. cerevisiae* using Human Pex3p and Pex19p. This project builds on previous work by constructing a human-yeast hybrid organelle *in vivo*. Human Pex19p has previously been shown to bind to yeast Pex3p (Fransen, Wylin et al. 2001; Fransen, Brees et al. 2002). Although the human Pex19p – yeast Pex3p interaction has previously been shown, we are unaware of any literature demonstrating the creation of peroxisomes functional in the import of PTS-tagged proteins using human Pex3p and Pex19p in yeast.

The analysis script output an Excel data file that was normalized and analyzed using Matlab. The result of this numerical analysis was a scatter heat map of the YFP vs. CFP peroxisome intensity and a histogram of the log of the CFP to YFP ratio. The curve fitting toolbox was used to fit distributions to this histogram to analyze the dynamics of the populations. The sum of the square of the residuals was calculated to assess the quality of the fitted distributions. The four tri-channel control strains were analyzed using Matlab and the scatter heat map and the distribution fit analysis were compared and discussed relative to the microscopy images. The control strains were KJY11, KJY14 and KJY15. From the control strains, KJY14 showed the most interesting results. The strain showed two populations in the scatter heat map and although the data were sparse, it showed a bimodal distribution in the histogram. When looking at the images to confirm this analysis, cytoplasmic localization can be clearly seen leading to these results being not a true representation of the actual situation.

The main experimental strain KJY10 showed a single population of peroxisomes on the normalized scatter heat map. The distribution that best fitted the histogram was a normal distribution, indicating a complete lack of bimodality. The microscopy images were then reviewed and by visual inspection showed no evidence of two populations of peroxisomes. This

experiment therefore produced no evidence of bimodality when examining the interactions of human and yeast Pex3p and Pex19p peroxins. The four proteins have been shown to cooperate to produce a single population of chimeric human-yeast peroxisomes. The project was successful in creating a yeast strain that produced peroxisomes in yeast using human Pex3p and Pex19p.

The author was also involved in continuing work on the arsenic biosensor first described during the iGEM competition (Aleksic, Bizzari et al. 2007). The continuation of this work involved designing high-throughput colorimetric pH assay system that was used to quantify arsenic concentrations in groundwater analogs. The publication, entitled "A pH-based biosensor for detection of arsenic in drinking water" was submitted, peer-reviewed and at the time of publication of this Thesis, was being prepared for re-submission. This publication can be found in Appendix 3: ABC paper on page 169. A complete list of the author's publications can be found on page 152.

7.4 Further Work

The promoter measurement work in *E. coli* could be continued by examining inducible promoters and promoters that use other sigma factors to affect transcription. A method would need to be designed so that standard measures of, for example IPTG, could be reliably and repeatedly used to induce promoters to the same level of activity. This could then be further developed as a method to measure the activity of promoters inducible by the presence of biotic or abiotic factors. This work could also be applied to other organisms such as *B. subtilis* or *S. elongatus*, where a reference promoter would need to be defined. Future promoter measurement work in *S. cerevisiae* could continue the experiments to demonstrate suitability of the *ADH1* promoter as the yeast reference promoter and establish a standard measurement protocol and media conditions. Another project could investigate finding in nature or designing a promoter that responds in a universal manner to all media conditions and investigating how to measure inducible promoters. This promoter would probably need to be engineered as it is unlikely to exist in current biology.

With the initial results indicating no bimodality, the door remains open to further experiments to create an artificial population of peroxisomes. The necessity of having two populations of

peroxisomes is a matter of debate as yeast could survive without endogenous peroxisomes, rendering bimodality in the peroxisome population unnecessary if synthetic peroxisomes are the intended outcome. A system could be developed to create temporary bimodality by putting key proteins such as Pex3p and Pex19p under control of inducible promoters. The *PEX* genes are induced using oleic acid and raffinose. Constructs could be made to express the key peroxins under the control of promoters induced by IPTG for a peroxisome on/off switch or constitutively expressed using the hexose transport (*HXT*) genes. These genes would include *PEX3* and *PEX19* but also a series of other genes.

Protein engineering would be a viable option to further the work on creating artificial organelles. If the binding domains of the human and yeast Pex3p-Pex19p interaction could be re-designed to be sufficiently different from each other, then bimodality might be observed. As the protein structures have now been resolved for Pex3p and Pex19p, this is now a more realistic possibility (Schmidt, Treiber et al. 2010; Schueller, Holton et al. 2010). This could be done by using binding domains from another species that have no similar motifs in yeast.

One option for further work with the current data-set would be to investigate the stochasticity regarding the number of peroxisomes per cell. The cell of origin information was captured, but not analysed so the data are present. Literature quotes cells as having one or two peroxisomes when the cells are growing under normal conditions (YEDP/YPDA). What makes a cell create a single peroxisome or several peroxisomes in a population of yeast cells? Is this behavior deterministic or random? In the context of the human and yeast peroxisome strain, does this behavior change from generation to generation and over time, does the population prioritize a specific species?

Following up on this question, the stochasticity could be investigated from generation to generation using time lapse fluorescence microscopy. An experiment could be setup to monitor the population over a period of 24 hours. Software would need to be written to track the cell of origin during the imaging and to assess the nature of the peroxisomes during the experiment. If a cell has solely yeast or human peroxisomes, what characterizes the daughter cell? Does it show the same characteristics as the mother cell, a homogenous population of peroxisomes or

the exact opposite? If these cells are tracked over time, is there any way to change the population composition by changing the media type or by modifying other external factors?

7.5 Perspective on Synthetic Biology

One of the greatest achievements in science was the sequencing of the human genome. Since the completion of the project, another 10 people have had their genomes sequenced and projects are in place to sequence hundreds and even thousands of individual genomes in an attempt to understand medicine at a genetic level. As sequencing capacity has increased exponentially, costs have fallen proportionately. Although DNA synthesis costs are still high relative to sequencing, the costs and error rates are eventually expected to fall, paving the way to a new world of digital biology. However, building new synthetic genes, proteins, pathways and eventually new organisms is not as simple as entering a genetic sequence into a web form and pressing the "order" button.

Many major challenges are still facing the field. One concerns part characterization and function. Parts can be defined as single genes, promoters, terminators, assemblies of parts and even entire metabolic networks could be built onto a single plasmid or integration vector. How are these parts characterized? Will libraries of parts be necessary for each organism? How will characterization differ across organisms? Another of the issues is standardization. How do you standardize protein function? How accurate do the measurements need to be? What are the tolerances on these measurements and how far out of the specification can these parts be before biological malfunction will crash the system? This information will likely be highly context specific; in terms of building a simple system to express GFP, the system will likely be robust to fluctuations in promoter expression levels. When engineering a metabolic pathway to produce pharmaceuticals or chemicals, flux of different intermediaries will be subject to much tighter control, both in terms of the robustness of the system to external perturbations and of the financial viability of the industrial process. More protein-protein interactions are also needed to further the field of Synthetic Biology. The ideal situation would be to write a coding region of DNA and be able to predict the functional structure. This problem has been described as being as difficult as accurately predicting the weather, so the challenge is not insignificant. Knowing the context in which the protein will function and other structures necessary for it to form will be key to this development, but in the mean time being able to reliably engineer functional protein

domains in constructs would be of immense value. There are parts for this purpose in the Registry of Standard Biological Parts, but they are generally not heavily used or well characterized.

These challenges are great, but are being addressed as the field of Synthetic Biology continues to develop. The field of biology is at a turning point. Progress made with single cell organisms will apply in other areas. Techniques will be developed to allow reliable specific integration of genetic material into mammalian cells. Genetic modifications will begin to move beyond the lab and into people's daily lives. A large part of the coming synthetic world of biology will involve cheaper, faster and better production of commercially valuable biological molecules such as drugs and fuels. Moving beyond commercial production, cells will be modified to fight cancer and cure human diseases in ways that pharmaceuticals will never be capable. The final stage of the synthetic world could be the advent of synthetic organisms, designed to solve problems far and beyond the capabilities of existing technology. This synthetic world will be ushered starting with small steps, designing custom organelles and measuring simple genetic regulatory function, as I have demonstrated in this thesis.

Publications

Peer reviewed publications:

de Mora, K., Joshi, N., Balint, B. L., Ward, F. B., Elfick, A., French, C.E.. *A pH-based biosensor for detection of arsenic in drinking water*, Anal Bioanal Chem, 2011. 400 p. 1031–1039.

Kelly, J. R.**,** Rubin, A. J.**,** Davis, J. H. **,** Ajo-Franklin, C. M.**,** Cumbers J.**,** Czar, M. J. , **de Mora, K.,** Glieberman, A. L.**,** Monie, D. D. and Endy, D., *Measuring the Activity of BioBrick Promoters Using an In Vivo Reference Standard.* J Biol Eng, 2009. **3**(1): p. 4.

Non-peer reviewed publications:

Aleksic J., Bizzari, F., Cai Y., Davidson, B., **de Mora, K.**, Ivakhno, S., Seshasayee, S., Nicholson, J., Wilson, J., Elfick, A., French, C. E., Kozma-Bognar, L., Ma, H., and Millar, A., *Development of a novel biosensor for the detection of arsenic in drinking water*, IET Synthetic Biology, 2007. **1**(2): p. 87-90.

Request for comments (RFC's):

Anderson, J. C., Che, A., Cowell, M., Elfick, A., **de Mora, K.**, Endy, D., French, C. E., Knight, T., Mayer, A., McArthur, G., Rettberg, R., Ridgway, D., Shetty, R., Sleight, S., and Tarjan D., *BBF RFC 0: Instructions to BBF RFC Authors*. MIT D-space, 2008.

References

Ajo-Franklin, C. M., D. A. Drubin, et al. (2007). "Rational design of memory in eukaryotic cells." Genes Dev **21**(18): 2271-6.

Alberti, S., A. D. Gitler, et al. (2007). "A suite of Gateway cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae." Yeast **24**(10): 913-9.

Alberts, B., A. Johnson, et al. (2002). Molecular Biology of the Cell. New York, Garland Science.

Aleksic, J., F. Bizzari, et al. (2007). "Development of a novel biosensor for the detection of arsenic in drinking water." IET Synth. Biol **1**((1-2)): 87-90.

Anderson, J. C. (2006). "BBa_J23101 constitutive promoter family." Retrieved 1-3-2008, from http://partsregistry.org/Part:BBa_J23101.

Andrianantoandro, E., S. Basu, et al. (2006). "Synthetic biology: new engineering rules for an emerging discipline." Mol Syst Biol **2**: 2006 0028.

Arkin, A. (2008). "Setting the standard in synthetic biology." Nature Biotechnology **26**(7): 771-4.

Arkin, A. P. (2001). "Synthetic cell biology." Current Opinion in Biotechnology **12**(6): 638-44.

Aslanidis, C. and P. J. de Jong (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." Nucleic Acids Res **18**(20): 6069-74.

Balmer, A. and P. Martin (2008). Synthetic Biology Social and Ethical Challenges. BBSRC. Nottingham, Andrew Balmer & Paul Martin.

Banerjee, S. K., P. S. Kessler, et al. (2005). "Identification of trypanosomatid PEX19: functional characterization reveals impact on cell growth and glycosome size and number." Mol Biochem Parasitol **142**(1): 47-55.

Beier, K. (1992). "Light microscopic morphometric analysis of peroxisomes by automatic image analysis: advantages of immunostaining over the alkaline DAB method." J Histochem Cytochem **40**(1): 115-21.

Beier, K. and H. D. Fahimi (1987). "Application of automatic image analysis for morphometric studies of peroxisomes stained cytochemically for catalase. II. Light-microscopic application." Cell Tissue Res **247**(1): 179-85.

Beier, K. and H. D. Fahimi (1992). "Application of automatic image analysis for quantitative morphological studies of peroxisomes in rat liver in conjunction with cytochemical staining with 3-3'-diaminobenzidine and immunocytochemistry." Microsc Res Tech **21**(4): 271-82.

Ben-Haim, N., P. Broz, et al. (2008). "Cell-specific integration of artificial organelles based on functionalized polymer vesicles." Nano Letters **8**(5): 1368-73.

Berg, P., D. Baltimore, et al. (1975). "Summary statement of the Asilomar conference on recombinant DNA molecules." Proc Natl Acad Sci U S A **72**(6): 1981-4.

Bhattachary, D., J. P. Calitz., et al. (2008). Synthetic Biology Dialogue**:** 89.

Bliss, M. (1982). The Discovery of Insulin. Chicago, University of Chicago Press.

Borst, P. (1989). "Peroxisome biogenesis revisited." Biochimica et Biophysica Acta **1008**(1): 1-13.

Brown, J. (2007). "The iGEM competition: building with biology." IET Synth. Biol **1**(1-2): 3-6.

Bugl, H., J. P. Danner, et al. (2007). "DNA synthesis and biological security." Nat Biotechnol **25**(6): 627-9.

Calvert, J. and P. Martin (2009). "The role of social scientists in synthetic biology. Science & Society Series on Convergence Research." EMBO Reports **10**(3): 201-4.

Canton, B., A. Labno, et al. (2008). "Refinement and standardization of synthetic biological parts and devices." Nature Biotechnology **26**(7): 787-93.

Carlson, R. H. (2010). Biology Is Technology: The Promise, Peril, and New Business of Engineering Life. Boston, Harvard University Press

Casadaban, M. J. (1975). "Fusion of the Escherichia coli lac genes to the ara promoter: a general technique using bacteriophage Mu-1 insertions." Proc Natl Acad Sci U S A **72**(3): 809-13.

Čeh, E., V. Kočar, et al. (2008). "Team Slovenia 2008 Home Page." from http://2008.igem.org/Team:Slovenia.

Chambers, A., J. S. Tsang, et al. (1989). "Transcriptional control of the Saccharomyces cerevisiae PGK gene by RAP1." Mol Cell Biol **9**(12): 5516-24.

Chen, M. (2007). "Building a New Organelle for Synthetic Biology." Retrieved August 2010, 2010, from http://parts.mit.edu/igem07/index.php/UCSF/Organelle_Intro.

Choi, C. Q. (2009). The First Synthetic Organelle. Scientific American.

Ciglič, M., O. Fekonja, et al. (2007). "Engineered human cells: say no to sepsis." IET Synth. Biol **1**(1-2): 13-16.

Ciglicˇ, M., O. Fekonja, et al. (2007). "Engineered human cells: say no to sepsis." IET Synth. Biol **1**(1-2): 13-16.

Cohen, S. N., A. C. Chang, et al. (1973). "Construction of biologically functional bacterial plasmids in vitro." Proc Natl Acad Sci U S A **70**(11): 3240-4.

Corish, P. and C. Tyler-Smith (1999). "Attenuation of green fluorescent protein half-life in mammalian cells." Protein Eng **12**(12): 1035-40.

De Duve, C. and P. Baudhuin (1966). "Peroxisomes (microbodies and related particles)." Physiol Rev **46**(2): 323-57.

Denis, C. L., J. Ferguson, et al. (1983). "mRNA levels for the fermentative alcohol dehydrogenase of Saccharomyces cerevisiae decrease upon growth on a nonfermentable carbon source." J Biol Chem **258**(2): 1165-71.

Durai, S., M. Mani, et al. (2005). "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells." Nucleic Acids Res **33**(18): 5978-90.

Ellis, T., X. Wang, et al. (2009). "Diversity-based, model-guided construction of synthetic gene networks with predicted functions." Nat Biotechnol **27**(5): 465-71.

Elowitz, M. B. and S. Leibler (2000). "A synthetic oscillatory network of transcriptional regulators." Nature **403**(6767): 335-8.

Endy, D. (2005). "Foundations for engineering biology." Nature **438**(7067): 449-53.

Endy, D. (2010). "Human Practices." Retrieved 1-8-2010, from http://synberc.org/humanpractices.

Endy, D., D. Grewal, et al. (2010). "The BioBrick Public Agreement v1a." Retrieved 1-2-2010, 2010, from http://hdl.handle.net/1721.1/50999.

Eustice, D. C., P. A. Feldman, et al. (1991). "A sensitive method for the detection of beta-galactosidase in transfected mammalian cells." Biotechniques **11**(6): 739-40, 742-3.

Fagarasanu, A., M. Fagarasanu, et al. (2007). "Maintaining peroxisome populations: a story of division and inheritance." Annu Rev Cell Dev Biol **23**: 321-44.

Fagarasanu, A., F. D. Mast, et al. (2010). "Molecular mechanisms of organelle inheritance: lessons from peroxisomes in yeast." Nat Rev Mol Cell Biol **11**(9): 644-54.

Fagarasanu, M., A. Fagarasanu, et al. (2006). "Sharing the wealth: peroxisome inheritance in budding yeast." Biochim Biophys Acta **1763**(12): 1669-77.

Fang, Y., J. C. Morrell, et al. (2004). "PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins." J Cell Biol **164**(6): 863-75.

Fleischmann, R. D., M. D. Adams, et al. (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science **269**(5223): 496-512.

Fransen, M., C. Brees, et al. (2002). "Analysis of mammalian peroxin interactions using a non-transcription-based bacterial two-hybrid assay." Mol Cell Proteomics **1**(3): 243-52.

Fransen, M., I. Vastiau, et al. (2005). "Analysis of human Pex19p's domain structure by pentapeptide scanning mutagenesis." J Mol Biol **346**(5): 1275-86.

Fransen, M., T. Wylin, et al. (2001). "Human pex19p binds peroxisomal integral membrane proteins at regions distinct from their sorting sequences." Mol Cell Biol **21**(13): 4413-24.

Frow, E. K. (2009). "A forum for 'doing society and genomics'. Science & Society Series on Convergence Research." EMBO Rep **10**(4): 318-21.

Fujiki, Y., Y. Matsuzono, et al. (2006). "Import of peroxisomal membrane proteins: the interplay of Pex3p- and Pex19p-mediated interactions." Biochim Biophys Acta **1763**(12): 1639-46.

Galarneau, A. Z. (2008). "A Life Turned Upside Down." Retrieved August 2010, 2010, from http://www.caedefensefund.org/press/BuffNewsGalarneau.pdf.

Gardner, T. S., C. R. Cantor, et al. (2000). "Construction of a genetic toggle switch in Escherichia coli." Nature **403**(6767): 339-42.

Gerhardt, C. (1853). "Untersuchungen über die wasserfreien organischen Säuren." Justus Liebigs Annalen der Chemie **87**(1): 57-84.

Geuze, H. J., J. L. Murk, et al. (2003). "Involvement of the endoplasmic reticulum in peroxisome formation." Mol Biol Cell **14**(7): 2900-7.

Ghaedi, K., M. Honsho, et al. (2000). "PEX3 is the causal gene responsible for peroxisome membrane assembly-defective Zellweger syndrome of complementation group G." Am J Hum Genet **67**(4): 976-81.

Giacomini, A., V. Corich, et al. (1992). "Experimental conditions may affect reproducibility of the beta-galactosidase assay." FEMS Microbiology Letters **79**(1-3): 87-90.

Gibson, D. G., G. A. Benders, et al. (2008). "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome." Science **319**(5867): 1215-20.

Gibson, D. G., J. I. Glass, et al. (2010). "Creation of a bacterial cell controlled by a chemically synthesized genome." Science **329**(5987): 52-6.

Gotte, K., W. Girzalsky, et al. (1998). "Pex19p, a farnesylated protein essential for peroxisome biogenesis." Mol Cell Biol **18**(1): 616-28.

Gould, S. J., G. A. Keller, et al. (1990). "Peroxisomal protein import is conserved between yeast, plants, insects and mammals." EMBO J **9**(1): 85-90.

Gregory, S. G., K. F. Barlow, et al. (2006). "The DNA sequence and biological annotation of human chromosome 1." Nature **441**(7091): 315-21.

Hanczyc, M. M. and J. W. Szostak (2004). "Replicating vesicles as models of primitive cell growth and division." Curr Opin Chem Biol **8**(6): 660-4.

Hartley, J. L., G. F. Temple, et al. (2000). "DNA cloning using in vitro site-specific recombination." Genome Res **10**(11): 1788-95.

Haseloff, J. and J. Ajioka (2009). "Synthetic biology: history, challenges and prospects." Journal of the Royal Society Interface **6 Suppl 4**: S389-91.

Heiland, I. and R. Erdmann (2005). "Biogenesis of peroxisomes. Topogenesis of the peroxisomal membrane and matrix proteins." FEBS J **272**(10): 2362-72.

Hessel, A. (2009). "Introduction to Synthetic Biology " Retrieved August 2010, 2010.

Hettema, E. H. and A. M. Motley (2009). "How peroxisomes multiply." J Cell Sci **122**(Pt 14): 2331-6.

Hirsch, R. (2005). "The strange case of Steve Kurtz: Critical Art Ensemble and the price of freedom." Retrieved July 2010, 2010, from http://findarticles.com/p/articles/mi_m2479/is_6_32/ai_n13820077/.

Hoepfner, D., D. Schildknegt, et al. (2005). "Contribution of the endoplasmic reticulum to peroxisome formation." Cell **122**(1): 85-95.

Hoepfner, D., M. van den Berg, et al. (2001). "A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in Saccharomyces cerevisiae." J Cell Biol **155**(6): 979-90.

Hogenboom, S., G. J. Romeijn, et al. (2002). "Absence of functional peroxisomes does not lead to deficiency of enzymes involved in cholesterol biosynthesis." J Lipid Res **43**(1): 90-8.

Hohfeld, J., M. Veenhuis, et al. (1991). "PAS3, a Saccharomyces cerevisiae gene encoding a peroxisomal integral membrane protein essential for peroxisome biogenesis." J Cell Biol **114**(6): 1167-78.

Huh, W. K., J. V. Falvo, et al. (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-91.

Ito, H., Y. Fukuda, et al. (1983). "Transformation of intact yeast cells treated with alkali cations." J Bacteriol **153**(1): 163-8.

Johnson, C. Y. (2010). "Start-up aims to sequence human genomes for $30, in just a few hours." Retrieved August 2010, 2010, from http://www.boston.com/yourtown/cambridge/articles/2010/06/07/start_up_aims_to_sequence_human_genomes_for_30_in_just_a_few_hours/.

Jones, J. M., J. C. Morrell, et al. (2004). "PEX19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins." J Cell Biol **164**(1): 57-67.

Karow, J. (2010). "As Illumina Cuts Price of Personal Sequencing Service, CEO Says Market Growth Hinges on Analysis." Retrieved August 2010, 2010, from http://www.genomeweb.com//node/942242?hq_e=el&hq_m=735028&hq_l=5&hq_v=10825f54c8.

Keasling, J. D. and H. Chou (2008). "Metabolic engineering delivers next-generation biofuels." Nat Biotechnol **26**(3): 298-9.

Kelly, J. R. (2008). "The BioBricks Foundation:Standards/Technical/Measurement/Promoter characterization experiment (FACS)." 2008, from http://openwetware.org/wiki/The_BioBricks_Foundation:Standards/Technical/Measurement/Promoter_characterization_experiment_%28FACS%29.

Kelly, J. R., A. J. Rubin, et al. (2009). "Measuring the activity of BioBrick promoters using an in vivo reference standard." J Biol Eng **3**: 4.

Kelly, J. R. C., B. Shetty, R. P. . (2010). "OpenWetWare." Retrieved 1-6-2010, from http://openwetware.org/wiki/Main_Page.

Kerppola, T. K. (2006). "Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells." Nat Protoc **1**(3): 1278-86.

Kim, P. K., R. T. Mullen, et al. (2006). "The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER." J Cell Biol **173**(4): 521-32.

Kitney, R. I. (2009). Synthetic Biology: scope, applications and implications. London, The Royal Academy of Engineering.

Kitney, R. I., P. S. Freemont, et al. (2007). "Engineering a molecular predation oscillator." IET Synth. Biol. **1**(1-2): 68-70.

Knight, T. (2003). "Idempotent Vector Design for Standard Assembly of Biobricks." MIT Synthetic Biology Working Group: 1-11.

Knight, T. (2007). "Request For Comments (RFCs)." Retrieved July 2010, from http://bbf.openwetware.org/RFC.html.

Kovacs, W. J. and S. Krisans (2003). "Cholesterol biosynthesis and regulation: role of peroxisomes." Adv Exp Med Biol **544**: 315-27.

Kovacs, W. J., L. M. Olivier, et al. (2002). "Central role of peroxisomes in isoprenoid biosynthesis." Prog Lipid Res **41**(5): 369-91.

Kuldell, N. and N. Lerner (2009). Genome Refactoring, Synthesis Lectures on Synthetic Biology. San Rafael, Morgan and Claypool publishers.

Kunau, W. H. (2005). "Peroxisome biogenesis: end of the debate." Curr Biol **15**(18): R774-6.

Kurbatova, E. M., T. A. Dutova, et al. (2005). "[Structural, functional and genetic aspects of peroxisome biogenesis]." Genetika **41**(2): 149-65.

Lauerman, J. (2009). "Complete Genomics Drives Down Cost of Genome Sequence to $5,000 " Retrieved August 2010, from http://www.bloomberg.com/apps/news?pid=newsarchive&refer=home&sid=aEUlnq6ltPpQ.

Lazarow, P. B. and Y. Fujiki (1985). "Biogenesis of peroxisomes." Annu Rev Cell Biol **1**: 489-530.

Levskaya, A., A. A. Chevalier, et al. (2005). "Synthetic biology: engineering Escherichia coli to see light." Nature **438**(7067): 441-2.

Levy, S., G. Sutton, et al. (2007). "The diploid genome sequence of an individual human." PLoS Biol **5**(10): e254.

Martin, J. G., M. Gupta, et al. (2009). "Toward an artificial Golgi: redesigning the biological activities of heparan sulfate on a digital microfluidic chip." J Am Chem Soc **131**(31): 11041-8.

Marzioch, M., R. Erdmann, et al. (1994). "PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes." EMBO J **13**(20): 4908-18.

Matsuzaki, T. and Y. Fujiki (2008). "The peroxisomal membrane protein import receptor Pex3p is directly transported to peroxisomes by a novel Pex19p- and Pex16p-dependent pathway." J Cell Biol **183**(7): 1275-86.

Matsuzono, Y., N. Kinoshita, et al. (1999). "Human PEX19: cDNA cloning by functional complementation, mutation analysis in a patient with Zellweger syndrome, and potential role in peroxisomal membrane assembly." Proc Natl Acad Sci U S A **96**(5): 2116-21.

McPherson, J. D., M. Marra, et al. (2001). "A physical map of the human genome." Nature **409**(6822): 934-41.

Miller, C. G. and D. Zipser (1977). "Degradation of Escherichia coli beta-galactosidase fragments in protease-deficient mutants of Salmonella typhimurium." J Bacteriol **130**(1): 347-53.

Miller, J. H. (1972). Experiments in molecular genetics. New York, Cold Spring Harbor, N.Y.

Miyata, N. and Y. Fujiki (2005). "Shuttling mechanism of peroxisome targeting signal type 1 receptor Pex5: ATP-independent import and ATP-dependent export." Mol Cell Biol **25**(24): 10822-32.

Monod, J. and M. Cohn (1952). "[Biosynthesis induced by enzymes; enzymatic adaptation.]." Adv Enzymol Relat Subj Biochem **13**: 67-119.

Motley, A. M. and E. H. Hettema (2007). "Yeast peroxisomes multiply by growth and division." J Cell Biol **178**(3): 399-410.

Mukai, S., K. Ghaedi, et al. (2002). "Intracellular localization, function, and dysfunction of the peroxisome-targeting signal type 2 receptor, Pex7p, in mammalian cells." J Biol Chem **277**(11): 9548-61.

Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harbor Symposia on Quantitative Biology **51 Pt 1**: 263-73.

Muntau, A. C., P. U. Mayerhofer, et al. (2000). "Defective peroxisome membrane synthesis due to mutations in human PEX3 causes Zellweger syndrome, complementation group G." Am J Hum Genet **67**(4): 967-75.

Murphy, R. F., M. V. Boland, et al. (2000). "Towards a systematics for protein subcelluar location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images." Proc Int Conf Intell Syst Mol Biol **8**: 251-9.

Niemisto, A., J. Selinummi, et al. (2006). "Extraction of the number of peroxisomes in yeast cells by automated image analysis." Conf Proc IEEE Eng Med Biol Soc **1**: 2353-6.

Ohtani, M., A. Saka, et al. (2004). "Development of image processing program for yeast cell morphology." J Bioinform Comput Biol **1**(4): 695-709.

Paciello, L., E. de Alteriis, et al. (2009). "Performance of the auxotrophic Saccharomyces cerevisiae BY4741 as host for the production of IL-1beta in aerated fed-batch reactor: role of ACA supplementation, strain viability, and maintenance energy." Microb Cell Fact **8**: 70.

Packham, E. A., I. R. Graham, et al. (1996). "The multifunctional transcription factors Abf1p, Rap1p and Reb1p are required for full transcriptional activation of the chromosomal PGK gene in Saccharomyces cerevisiae." Mol Gen Genet **250**(3): 348-56.

Parkinson, J., G. Cromar, et al. (2009). "Toronto MaRS Discovery iGEM 2009 project." Retrieved August 2010, from http://2009.igem.org/Team:TorontoMaRSDiscovery/Team.

Paterson, L., S. King, et al. (2009). Synthetic Biology: public dialogue on synthetic biology. London, The Royal Academy of Engineering**:** 60.

Phillips, I. E. and P. A. Silver (2006). "A New Biobrick Assembly Strategy Designed for Facile Protein Engineering." DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group.

Pontikis, A. (2010). "The Nobel Prize in Physiology or Medicine 1945: Sir Alexander Fleming, Ernst B. Chain, Sir Howard Florey." Nobel Prize Retrieved August 2010, from http://nobelprize.org/nobel_prizes/medicine/laureates/1945/fleming.html.

Purdue, P. E. and P. B. Lazarow (2001). "Peroxisome biogenesis." Annu Rev Cell Dev Biol **17**: 701-52.

Randerson, J. (2006). "Revealed: the lax laws that could allow assembly of deadly virus DNA." Retrieved September 2008, 2008, from http://www.guardian.co.uk/world/2006/jun/14/terrorism.topstories3.

Reifenberger, E., K. Freidel, et al. (1995). "Identification of novel HXT genes in Saccharomyces cerevisiae reveals the impact of individual hexose transporters on glycolytic flux." Mol Microbiol **16**(1): 157-67.

Rettberg, R. (2006). "BioBrick Assembly: Standard assembly." Retrieved July 2007, from http://partsregistry.org/Assembly:Standard_assembly.

Rettberg, R. (2010). "BioBrick Documentation Help Page." Registry of Standard Biological Parts help pages Retrieved 01-07-2010, from http://partsregistry.org/Help:Contents.

Rettberg, R. (2010). "iGEM 2010." Retrieved 1-7-2010, from http://2010.igem.org/Main_Page.

Rettberg, R. (2010). "The Registry of Standard Biological Parts." Retrieved August 2010, 2010, from http://partsregistry.org/Main_Page.

Rettberg, R. (2010). "Registry of Standard Biological Parts Catalog." Retrieved 1-6-2010, from http://partsregistry.org/Catalog.

Rhodin, J. (1954). Correlation of ultrastructural organization and function in normal and experimentally changed proximal convoluted tubule cells of the mouse kidney. Aktiebolaget Godvil. Stockholm. **Ph. D.**

Ro, D. K., E. M. Paradise, et al. (2006). "Production of the antimalarial drug precursor artemisinic acid in engineered yeast." Nature **440**(7086): 940-3.

Roodbeen, R. and J. C. van Hest (2009). "Synthetic cells and organelles: compartmentalization strategies." Bioessays **31**(12): 1299-308.

Rottensteiner, H., A. Kramer, et al. (2004). "Peroxisomal membrane proteins contain common Pex19p-binding sites that are an integral part of their targeting signals." Mol Biol Cell **15**(7): 3406-17.

Sacksteder, K. A., J. M. Jones, et al. (2000). "PEX19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis." J Cell Biol **148**(5): 931-44.

Salis, H. M., E. A. Mirsky, et al. (2009). "Automated design of synthetic ribosome binding sites to control protein expression." Nat Biotechnol **27**(10): 946-50.

Sambrook, J., E. Fritsch, et al. (1989). Molecular cloning: a laboratory manual. New York, Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5463-7.

Sasaki, Y., T. Sone, et al. (2004). "Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system." J Biotechnol **107**(3): 233-43.

Sato, Y., H. Shibata, et al. (2008). "Characterization of the interaction between recombinant human peroxin Pex3p and Pex19p: identification of TRP-104 IN Pex3p as a critical residue for the interaction." Journal of Biological Chemistry **283**(10): 6136-44.

Schafer, H., K. Nau, et al. (2001). "Identification of peroxisomal membrane proteins of Saccharomyces cerevisiae by mass spectrometry." Electrophoresis **22**(14): 2955-68.

Schmidt, F., N. Treiber, et al. (2010). "Insights into peroxisome function from the structure of PEX3 in complex with a soluble fragment of PEX19." J Biol Chem **285**(33): 25410-7.

Schueller, N., S. J. Holton, et al. (2010). "The peroxisomal receptor Pex19p forms a helical mPTS recognition domain." EMBO J **29**(15): 2491-500.

Serebriiskii, I. G. and E. A. Golemis (2000). "Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system." Anal Biochem **285**(1): 1-15.

Sharma, K. G., R. Kaur, et al. (2003). "The glutathione-mediated detoxification pathway in yeast: an analysis using the red pigment that accumulates in certain adenine biosynthetic mutants of yeasts reveals the involvement of novel genes." Arch Microbiol **180**(2): 108-17.

Sikorski, R. S. and P. Hieter (1989). "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae." Genetics **122**(1): 19-27.

SKF. (2010). "SKF Group." Retrieved 1-7-2010, from http://www.skf.com/portal/skf/home?lang=en&site=COM.

Smith, H. O. and K. W. Wilcox (1970). "A restriction enzyme from Hemophilus influenzae. I. Purification and general properties." J Mol Biol **51**(2): 379-91.

Smolke, C. D. (2009). "Building outside of the box: iGEM and the BioBricks Foundation." Nat Biotechnol **27**(12): 1099-102.

Soukupova, M., C. Sprenger, et al. (1999). "Identification and characterization of the human peroxin PEX3." Eur J Cell Biol **78**(6): 357-74.

South, S. T. and S. J. Gould (1999). "Peroxisome synthesis in the absence of preexisting peroxisomes." J Cell Biol **144**(2): 255-66.

Stano, P. and P. L. Luisi (2010). "Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells." Chem Commun (Camb) **46**(21): 3639-53.

Swain, P. S., M. B. Elowitz, et al. (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proc Natl Acad Sci U S A **99**(20): 12795-800.

Sweet, D. J. (2010). ASSOCIATION FOR MOLECULAR PATHOLOGY, ET AL. against UNITED STATES PATENT AND TRADEMARK OFFICE, ET AL. New York, UNITED STATES DISTRICT COURT SOUTHERN DISTRICT OF NEW YORK.

Szybalski, W. and A. Skalka (1978). "Nobel prizes and restriction enzymes." Gene **4**(3): 181-2.

Tabak, H. F., D. Hoepfner, et al. (2006). "Formation of peroxisomes: present and past." Biochim Biophys Acta **1763**(12): 1647-54.

Tam, Y. Y., A. Fagarasanu, et al. (2005). "Pex3p initiates the formation of a preperoxisomal compartment from a subdomain of the endoplasmic reticulum in Saccharomyces cerevisiae." J Biol Chem **280**(41): 34933-9.

Titorenko, V. I. and R. A. Rachubinski (2001). "The life cycle of the peroxisome." Nat Rev Mol Cell Biol **2**(5): 357-68.

Tsien, R. Y. (1998). "The green fluorescent protein." Annu Rev Biochem **67**: 509-44.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Walde, P., R. Wick, et al. (1994). "Autopoietic Self-Reproduction of Fatty Acid Vesicles." J. Am. Chem. SOC. **116**: 11649-11654.

Wanders, R. J. (2004). "Metabolic and molecular basis of peroxisomal disorders: a review." Am J Med Genet A **126A**(4): 355-75.

Wanders, R. J. and H. R. Waterham (2006). "Biochemistry of mammalian peroxisomes revisited." Annual Review of Biochemistry **75**: 295-332.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature **171**(4356): 737-8.

Weisman, L. S., R. Bacallao, et al. (1987). "Multiple methods of visualizing the yeast vacuole permit evaluation of its morphology and inheritance during the cell cycle." J Cell Biol **105**(4): 1539-47.

Weller, S., S. J. Gould, et al. (2003). "Peroxisome biogenesis disorders." Annual Review of Genomics & Human Genetics **4**: 165-211.

Wieden, H., J. Fischer, et al. (2009). "University of Lethbridge iGEM 2009 Project Overview." iGEM 2009 Projects Retrieved August 2010, from http://2009.igem.org/Team:Lethbridge.

Win, M. N. and C. D. Smolke (2007). "A modular and extensible RNA-based gene-regulatory platform for engineering cellular function." Proc Natl Acad Sci U S A **104**(36): 14283-8.

Wöhler, F. (1828). "Poggendorff's Ann." Phys. **12**: 253-256.

Yeh, B. J. and W. A. Lim (2007). "Synthetic biology: lessons from the history of synthetic organic chemistry." Nature Chemical Biology **3**(9): 521-5.

Zonneveld, B. J. and A. L. van der Zanden (1995). "The red ade mutants of Kluyveromyces lactis and their classification by complementation with cloned ADE1 or ADE2 genes from Saccharomyces cerevisiae." Yeast **11**(9): 823-7.

Appendix 1: BBa_F2620 datasheet

BBa_F2620

$3OC_6HSL \rightarrow$ PoPS Receiver

http://parts.mit.edu/registry/index.php/Part:BBa_F2620

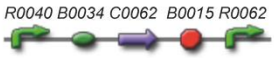
Authors:
Barry Canton [bcanton@mit.edu]
Anna Labno [labnoa@mit.edu]

Last Update: 19 October 2007

Description

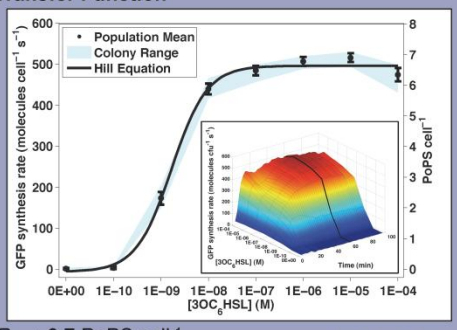
A transcription factor (LuxR, BBa_C0062) that is active in the presence of cell-cell signaling molecule 3OC6HSL is controlled by a TetR-regulated operator (BBa_R0040). Device input is 3OC6HSL. Device output is PoPS from a LuxR-regulated operator. If used in a cell containing TetR then a second input signal such as aTc can be used to produce a Boolean AND function.

Parts



BBa_C0062: luxR ORF
BBa_R0040: LuxR-regulated operator
BBa_R0040: TetR-regulated operator

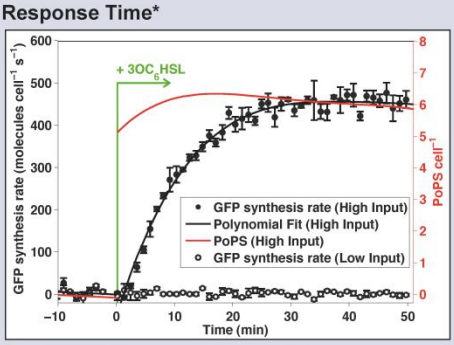
Transfer Function*



P_{max}: **6.7** PoPS cell^{-1}
K: **2E-09** M $3OC_6HSL$
n: **1.2**

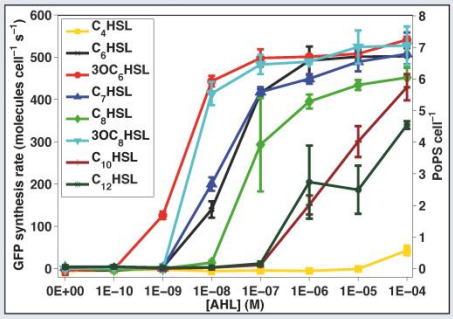
$$P_{out} = P_{max} \frac{[3OC_6HSL]^n}{K^n + [3OC_6HSL]^n}$$

Response Time*



BBa_F2620 Response Time: **<1** *min*
(PoPS calculated from polynomial fit to GFP synthesis rate data. High/Low input - 1E-7/0 M $3OC_6HSL$)

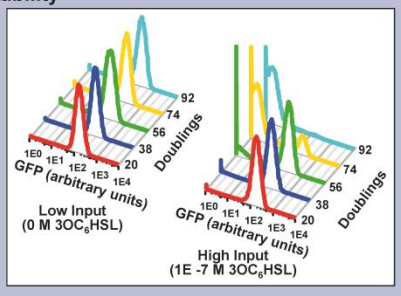
Input Compatibility*



Part Compatibility

Chassis: *MC4100*, *MG1655*, and *DH5α*
Plasmids: *pSB3K3* and *pSB1A2*
Devices: *E0240*, *E0430* and *E0434*
Crosstalk with systems containing *C0040*

Stability**



Genetic: **>92/<74** replication events**
Performance: **>92/<74** replication events**
(low/high input)

Transcriptional Output Demand (low/high input)
Nucleotides: **0.2xNt / 6xNt** nucleotides cell^{-1} s^{-1}
Polymerases: **4.4E-3xNt / 1.5E-1xNt** RNAP cell^{-1}
(**Nt** = downstream transcript length)

Conditions (abridged)
Output: PoPS measured via *BBa_E0240*
Culture: Supplemented M9, 37ºC
Vector: pSB3K3
Chassis: MG1655
**Equipment:* PE Victor3 plate reader
***Equipment:* BD FACScan cytometer

Registry of Standard Biological Parts
making life better, one part at a time

Signaling Devices

Appendix 2: Supplemental Protocols

Methods Specific to the E. coli Promoter Measurement Project

A cell growth and measurement protocol was to suit the multi-institution measurements. The aim of the protocol design was to establish something easy enough to allow an iGEM team to take the measurements with little or no specialized training. No attempts were made to standardize the equipment across the testing institutions.

E. coli Culture Protocol

The multi-institution protocol was developed and tested using a flow cytometer (Kelly 2008):

1) Streak LB + Kan plates of the 5 test strains and an LB (no antibiotic) for TOP10
2) For each test construct add 5ml of supplemented M9-glycerol medium and Kanamycin (20 µg/ml) to three 14 ml test tubes (15 tubes total). Add 5 ml of supplemented M9 medium (w/glycerol) with no antibiotic for the TOP10 negative control cells to three 14 ml test tubes.
3) Inoculate media by picking single colonies from plates using tips. Grow cultures for 20 hrs at 37 °C with aeration at 200 rpm.
4) In the morning dilute 100x into 5ml of pre-warmed (37 ^{o}C) fresh media to get cells back into log phase and let grow for 4hrs under same conditions as overnight
5) The cultures should be at an OD of 0.2 - 0.4 after 4hrs.

Live Cell E. coli Flow Cytometry Procedure

Cells were collected at two time points while the cultures were in exponential phase. The GFP channel was measured once the flow cytometer had been calibrated (Kelly 2008):

1) When the 5 ml culture reaches OD600 ~ 0.300, pellet the cells by centrifuging 1ml of each culture in an Eppendorf at 3000 g at room temperature for 5 minutes.
2) Remove the supernatant.
3) Resuspend the cell pellet in 1 ml PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Sodium Phosphate dibasic, 2 mM Potassium Phosphate monobasic, pH 7.4).

4) Add 250 µl washed cells through the cell strainer lid into a 5 ml polystyrene tube. To get the solution to pass through the strainer, apply slight pressure to the strainer lid with the pipette tip as dispensing the cell solution.
5) Add 250 µl PBS through the cell strainer lid into the 5 ml polystyrene tube.
6) Place cells on ice.
7) Analyze cells by flow cytometry using 488 nm excitation as quickly as possible.

S. cerevisiae growth protocol

This protocol was developed for use with the BD LSD-II flow cytometer as used with the *E. coli* promoter measurement experiments. It was adapted to work with yeast, and with a greater number of samples.

1) Fill a 96 well deep-well plate with 750 µl of YEPD media per well
2) Inoculate using a 96 head pin tool from the frozen glycerol stock plate
3) Grow the plate overnight to saturation in a shaking incubator at 30 ^{o}C and 1000 rpm. The high rpm was achieved in an Appropriate Technical Resources (ATR) Multitron 2
4) Fill a 96 well U-bottom plate with 200 µl of SD or S-KAc media per well
5) Inoculate the 96 well deep-well plate using 50 µl of culture (to an OD of ~0.1 at 600 nm) and place in shaking incubator for 3-4 h
6) Spin down the plate and extract media with vacuum line/multi-head pipette
7) Wash cells using ddH_20 twice
8) Re-suspend cells in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Sodium Phosphate dibasic, 2 mM Potassium Phosphate monobasic, pH 7.4), put on ice and analyze as soon as possible

Appendix 3: Publication in the Journal Analytical and Bioanalytical Chemistry

K. de Mora, N. Joshi, B. L. Balint, F. B. Ward, A. Elfick & C. E. French

Anal Bioanal Chem (2011) 400:1031–1039
DOI 10.1007/s00216-011-4815-8

ORIGINAL PAPER

A pH-based biosensor for detection of arsenic in drinking water

K. de Mora · N. Joshi · B. L. Balint · F. B. Ward · A. Elfick · C. E. French

Received: 13 October 2010 / Revised: 13 February 2011 / Accepted: 16 February 2011 / Published online: 27 March 2011

Abstract Arsenic contaminated groundwater is estimated to affect over 100 million people worldwide, with Bangladesh and West Bengal being among the worst affected regions. A simple, cheap, accurate and disposable device is required for arsenic field testing. We have previously described a novel biosensor for arsenic in which the output is a change in pH, which can be detected visually as a colour change by the use of a pH indicator. Here, we present an improved formulation allowing sensitive and accurate detection of less than 10 ppb arsenate with static overnight incubation. Furthermore, we describe a cheap and simple high-throughput system for simultaneous monitoring of pH in multiple assays over time. Up to 50 samples can be monitored continuously over the desired time period. Cells can be stored and distributed in either air-dried or freeze-dried form. This system was successfully tested on arsenic-contaminated groundwater samples from the South East region of Hungary. We hope to continue to develop this sensor to produce a device suitable for field trials.

Keywords Biosensors · Groundwater · Arsenic · Assay · BioBricks

Published in the special issue *Microorganisms for Analysis* with Guest Editor Gérald Thouand.

K. de Mora · A. Elfick
School of Engineering, University of Edinburgh,
Edinburgh EH9 3JL, UK

N. Joshi
School of GeoSciences, University of Edinburgh,
Edinburgh EH9 3JW, UK

B. L. Balint
Department of Biochemistry and Molecular Biology,
University of Debrecen Medical and Health Science Centre,
Debrecen H-4032, Hungary

F. B. Ward · C. E. French
School of Biological Sciences, University of Edinburgh,
Edinburgh EH9 3JR, UK

C. E. French (✉)
University of Edinburgh,
Darwin Building, King's Buildings,
Edinburgh EH9 3JR, UK
e-mail: C.French@ed.ac.uk

Introduction

Arsenic in drinking water, in the form of arsenate or arsenite anions, is a major public health issue in a number of regions worldwide and is prevalent in South and South East Asia, especially in Bangladesh and West Bengal [1, 2]. Recent studies have shown that arsenic poisoning may be worse than previously thought, with as many as 77 million people possibly exposed to arsenic contaminated water in Bangladesh alone [3]. Studies have also shown groundwater contaminated with arsenic at concentrations up to 500 ppb in one of India's most heavily industrialised regions: the Thane region [4]. Arsenic is immediately toxic in high concentrations (~60,000 ppb), but exhibits chronic effects at lower concentrations [5]. The World Health Organization has set a maximum safe limit of 10 ppb (10 $\mu g\ l^{-1}$), but many regions still consider 50 ppb (50 $\mu g\ l^{-1}$) to be a safe target at the present time. Symptoms of inorganic arsenic poisoning can occur at concentrations varying from 300–30,000 ppb and include vomiting, stomach and intestinal irritation, nausea and diarrhoea [5]. One of the most visible signs of arsenic poisoning is skin lesions that can appear on the hands, feet and torso. These lesions can eventually result in skin cancer if untreated, but more often lead to social ostracism as arsenicosis is perceived

to be a contagious disease [6]. At lower concentrations, visible symptoms can take months or years to appear.

Current testing in the affected regions is most widely conducted using atomic adsorption spectroscopy (M. Owens, personal communication). This technique is quantitative and reliable [7] but problematic, as water samples must be shipped from contaminated sites to testing laboratories. The capital investment per system is in the order of 20,000 USD, with an additional cost for consumables and the use of a qualified technician. Various methods also exist to test for arsenic in the field, most notably the Wagtech Arsenator and paper test strips based on the Gutzeit method. The Arsenator produces a quantitative readout of the arsenic concentration based on an electronic read of the Gutzeit method [8]. The basic mechanism behind the Gutzeit-based field test strips is conversion of arsenic compounds into arsenic trihydride by zinc. The hydride then stains paper impregnated with mercuric bromide [9].

Although assays based on the Gutzeit method have been deployed in the past, researchers are looking towards enzymatic and biological approaches to detect arsenic. One approach is to use screen-printed electrodes (SPCE) with acetylcholinesterase (Ach) and to measure an amperometric response. Sanllorente-Mendez et al. have demonstrated an approach where they can detect arsenic to a limit of 1.1×10^{-8} M for their Ach/SPCE biosensor [10]. Although successful, they reported storing their electrodes at 4 °C, which may make this device difficult to transport and store in environments lacking refrigeration. Other groups have reported using the electrostatic oxidation of L-cysteine on screen-printed electrodes to determine arsenate concentration. Sarkar et al. immobilised L-cysteine on a working electrode by in situ polymerization of acrylamide and determined that their system produced a linear response to arsenic from below 1 to 30 ppb [11]. This work showed functional and sensitive arsenic detection using screen-printed electrodes; however, the most sensitive platinum variants may not be cost-efficient in terms of developing a field test device.

Whole-cell biosensors offer a potential alternative method for arsenic detection [12]. Previous arsenic bioassay systems have been developed using bioluminescence, Green Fluorescent Protein, and β-galactosidase as the reporter mechanisms [13, 14]. Previous work from van der Meer and colleagues demonstrated a system combining whole cell *Escherichia coli* biosensors with a microfluidics environment allowing arsenic concentration measurements to be taken from as few as 200 cells [15]. Bacteriological arsenic biosensors have been also been previously field tested [16]. We have previously described a biosensor based on genetically modified *E. coli* cells in which the presence of arsenate or arsenite induces expression of β-galactosidase, allowing fermentation of lactose and consequent change in pH which can be read using a simple pH indicator [17, 18]. Here, we describe further characterization of the strain, an improved formulation of assay medium that allows for static incubation of the samples and a new assay allowing continuous monitoring of up to 50 samples simultaneously.

Materials and methods

Organisms and growth conditions

Construction of our arsenic biosensor organism, *E. coli* JM109/pSB1A2-BBa_J33203, has been previously described [17]. The BioBrick assay construct, pSB1A2-BBa_J33203, is available from the Registry of Standard Biological Parts, hosted at Massachusetts Institute of Technology, as are the subcomponents BBa_J33201 (arsenic-responsive promoter) and BBa_J33202 (*lacZ'* reporter gene) used to generate it. Initial experiments were conducted in Luria-Bertani (LB) broth with 10 g/l lactose and 100 μg/ml ampicillin. Later experiments used a variant of Hugh and Leifson's OF (oxidation–fermentation) medium designated Arsenic Biosensor Medium (ABM) 6. ABM6 contains 2.0 peptone, 0.1 yeast extract, 0.3 K_2HPO_4, 2.1 $NaHCO_3$, and 0.1 g/l bromothymol blue. Lactose (10 g/l) was added if required, but in most cases was present in the dried cells used as inoculum as described below. All components were mixed prior to autoclaving. Arsenic was added as sodium arsenate from sterilised stock solutions at 1,000 or 10,000 ppb arsenic. Variants of ABM6 media were made with 1,500 mg l^{-1} of iron (II) sulphate (527 mg l^{-1} iron) and 2 mg l^{-1} zinc sulphate (0.76 mg l^{-1} zinc) concentrations, respectively, representing nine and four times the highest reported concentrations in the British Geological Survey of Bangladeshi groundwater [19]. For preparation of air-dried cells, overnight cultures were grown in LB, mixed with an equal volume of 20% *w/v* lactose, dispensed at the appropriate volume into 1.5 ml microcentrifuge tubes, and incubated with open lids at 37 °C overnight, during which the cell–lactose mixture dried to a glassy state. For freeze-drying, overnight cultures were harvested by centrifugation, resuspended in 20% *w/v* sterile lactose solution, dispensed into 1.5 ml microcentrifuge tubes, frozen at −80 °C and dried in an Edwards Modulyo freeze drier for 15 h. Dried cells were stored at room temperature.

Assay procedures

In initial experiments, assays were conducted as described previously [17, 18], with 6 ml LB plus lactose and other additives in 20 ml sealed glass vials, incubated with shaking at 37 °C, and pH was measured at intervals using

a standard semi-micro glass pH electrode. In later assays, as specified, 1 ml ABM6 was added to a tube of dried cells, and the desired concentration of sodium arsenate was added from a stock solution. Unless otherwise specified, tubes were then incubated at 37 °C without shaking. For simultaneous visual monitoring of multiple assays, up to 50 tubes were placed in a specially constructed rack in a 37 °C incubator with a clear front cover. Images were captured using a Creative Live! Vista IM webcam (model no: VF0260) and Nimisis Flix time-lapse software was used for image acquisition. Jpeg images were acquired due to the small size of the files and the .AVI time-lapse videos were made with no file compression in Flix. Due to file header and footer issues, the .AVI videos were opened and re-saved in ImageJ before being processed in Image Pro 7.0. Once parsed in Image Pro 7.0, data were processed into Excel readable spreadsheets using Mathworks Matlab.

Miller assays for β-galactosidase activity were conducted according to a standard protocol [20]. Activity of XylE (catechol-2,3-dioxygenase) was monitored by adding catechol to a final concentration of 0.1 mM to cell suspension and monitoring absorbance of the yellow reaction product (2-hydroxy-*cis*,*cis*-muconic semialdehyde) at 420 nm.

Results

Effect of medium and inoculum level

Previously, we have described assays in continuously shaken LB cultures inoculated from a previous liquid culture, with pH monitored at intervals using a pH electrode [18]. Clearly, this assay configuration is not ideal for field use. We therefore sought to develop a system with more suitable characteristics for use by non-experts under field conditions with minimal laboratory equipment. To clarify the effect of different growth parameters, initially, we investigated the effect of using a more dilute growth medium. Use of LB diluted by a factor of 2 or 4 showed the higher dilution factor medium led to less pH decrease in arsenic-free controls while maintaining a strong pH response to 60 ppb arsenic, but further dilution led to a decreased pH response in the presence of arsenic. Use of different levels of inoculum was also investigated. A larger inoculum led to more rapid response, but where samples were incubated for longer periods, the level of initial inoculum made little difference to the final pH over the range tested (data not shown).

Effect of bicarbonate

In our previous characterization tests, we sought to determine the effect of common groundwater buffer ions on pH response [18]. Phosphate at relevant levels was found to have no effect on the assay but bicarbonate was unexpectedly found to apparently increase the sensitivity of the assay, leading to enhanced pH response at low arsenic conditions. To determine the basis for this effect, we constructed a modified version of the biosensor by adding a second reporter gene, *xylE* (encoding catechol-2,3-dioxygenase; BioBrick BBa_J33204) to the BioBrick biosensor plasmid, giving the new biosensor construct pSB1A2-BBa_J15501. Biosensor organisms were incubated in the presence of differing levels of bicarbonate and arsenate and were assayed for pH, β-galactosidase activity (Miller assay) and XylE activity. As previously observed, the pH assay showed an enhanced response to low arsenate concentrations in the presence of bicarbonate, whereas no difference was seen at higher arsenate concentrations (Fig. 1a); thus, the presence of bicarbonate appeared to increase the sensitivity of the assay. However, Miller assays showed that β-galactosidase activity was increased in the presence of bicarbonate at both low and high arsenate concentrations (Fig. 1b); thus, it appears that the apparent lack of effect of bicarbonate on the pH-based assay at high arsenate concentrations is an artefact caused by saturation of the pH response above a certain β-galactosidase activity. Furthermore, despite a high background activity, XylE activity was also seen to increase in the presence of arsenate (Fig. 1c), suggesting that the effect was due to increased induction of the reporter genes rather than some direct effect on the activity of β-galactosidase. It therefore seems likely that bicarbonate is increasing the interaction of arsenate with the ArsR repressor protein, perhaps by increasing entry of arsenate into the cell or modifying arsenate speciation. Further experiments may clarify this. For practical purposes, our previously described experiments [18] have indicated that the effect of bicarbonate appears to saturate at a concentration below 10 mM bicarbonate; thus, addition of sodium bicarbonate to all assays at a concentration above this should both enhance sensitivity and reduce variability due to varying bicarbonate concentrations in water samples.

Development of an assay medium suitable for static incubation

The effect of different incubation conditions on assays in LB was also investigated, using a variety of different containers and liquid–air ratios (data not shown). The most significant result was that in unshaken cultures, a large pH drop always occurred in arsenic-free controls, presumably due to the decreased aeration, making the test unreliable. Since we aim to develop an assay suitable for field use, this led us to investigate alternative formulations of assay medium. The OF (oxidation–fermentation) test is a standard

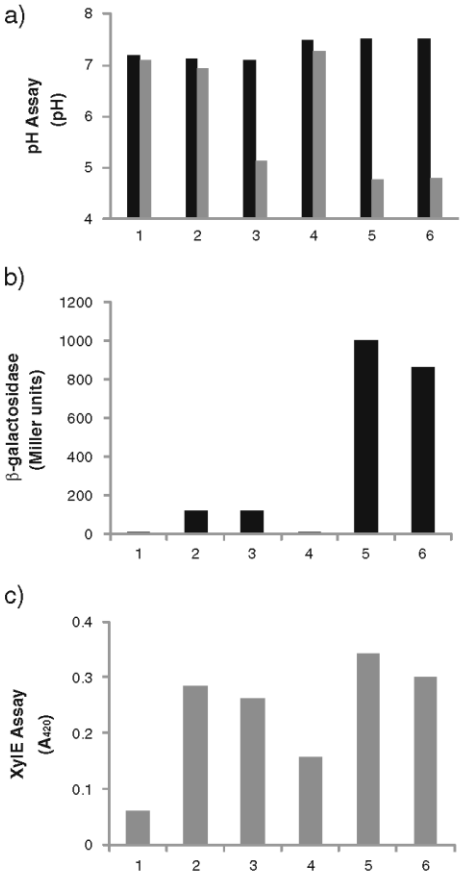


Fig. 1 Effect of bicarbonate on reporter activity. **a** Initial and overnight pH, **b** β-galactosidase (Miller assay), **c** XylE (catechol-2,3-dioxygenase) assay. Sample identification: *1* no bicarbonate, 0 ppb arsenic as arsenate. *2* no bicarbonate, 7.5 ppb arsenic as arsenate. *3* no bicarbonate, 15 ppb arsenic as arsenate. *4* 12.5 mM bicarbonate, 0 ppb arsenic as arsenate. *5* 12.5 mM bicarbonate, 7.5 ppb arsenic as arsenate. *6* 12.5 mM bicarbonate, 15 ppb arsenic as arsenate

and widely used test for the production of acid from carbohydrates and is commonly used in the identification of medically important bacteria [21]. This assay is normally conducted without shaking and uses the pH indicator bromothymol blue to generate a visual response. Various modifications of OF medium, with the addition of bicarbonate, were tested, eventually resulting in the development of ABM6 (see Materials and methods for composition), which was found to give a good colour response in less than 24 h in the presence of arsenate concentrations as low as 5 ppb arsenic but not to change colour over 24 h in the absence of arsenic, provided that a very low inoculum level was used (data not shown). This medium was used as the basis for further experiments.

Drying of cells for storage and distribution

Development of this assay for field use will require that cells be dried for storage and distribution. We have previously reported that lactose is a suitable cryoprotectant for freeze-drying of cells, resulting in as high as 30% survival under the conditions tested. This also provides lactose required for the assay. Since our previous experiments had shown that a small inoculum provides the best results for overnight assays under unshaken conditions, we investigated the possibility of air-drying as an alternative to freeze-drying, as this would remove the requirement for an expensive piece of equipment. Cells (initial volume of 20 to 100 μl per 1.5 ml microcentrifuge tube) were air-dried at 37 °C from a suspension containing 10% *w/v* lactose and were found to dry overnight to a glassy consistency. Figure 2 shows survival of cells dried from initial volumes of 20 or 100 μl of cell suspension (Fig. 2). Thus, the tubes containing 100 μl of cells initially showed higher survival but later showed more rapid die-off. This may be simply due to less effective drying leading to higher moisture content when sealed. Nevertheless, tubes of biosensor cells dried in this way showed reliable response to arsenic for at least 2 weeks after drying. Unless this can be improved with further development, freeze-drying will probably be required for longer-term storage.

Development of a colorimetric assay for the high-throughput quantification of pH measurements

To facilitate characterization of our biosensor system, we have developed a technique to measure the pH of multiple samples simultaneously over any desired period with high temporal resolution (Materials and methods). In order to calibrate the image acquisition system, we measured the absorbance of water samples with bromothymol blue adjusted to pH values in the range 10.91 to 3.24 (Fig. 3). The greatest change in absorbance was found to occur at 625 nm.

In the multi-tube system, assays are conducted by inoculating a series of 1.5 ml microcentrifuge tubes with ABM6 medium supplemented with arsenate, then placing them on a rack in a 37°C incubator in front of a Web camera. Images are acquired over the period of the experiment and compiled into a time lapse video which is read into Image Pro 7.0 and analysed using a custom-designed script. As the colour of the blue tubes stands out

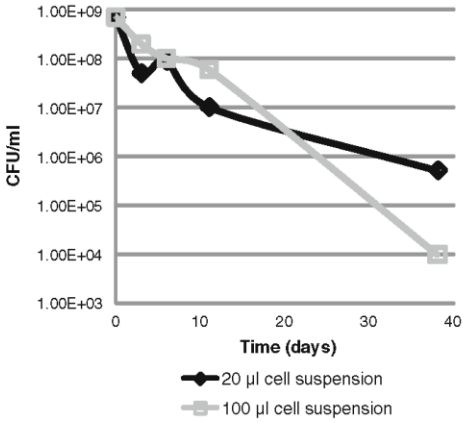


Fig. 2 Survival of air-dried cells following storage at room temperature

relative to the background, a colour threshold is applied to create regions of interest. For each region of interest, the red, green and blue colour channel intensity levels are averaged over the area of the tube and recorded in an Excel spreadsheet. The script applies the same regions of interest to each frame in the time-lapse creating RGB numerical values for each tube at each time step. The Excel file is read into MathWorks Matlab to parse the data into a three-dimensional matrix of the red, green and blue colour intensities. Initial examination of the raw data confirmed that the greatest colour change occurred on the red channel as seen at 625 nm in Fig. 3. Plots showing the colour change over time with bromothymol blue therefore used data from the red channel.

Calibration of the pH scale in the plot was achieved by imaging pH-adjusted samples of bromothymol blue. The suppliers of bromothymol blue (Sigma) specify a colour

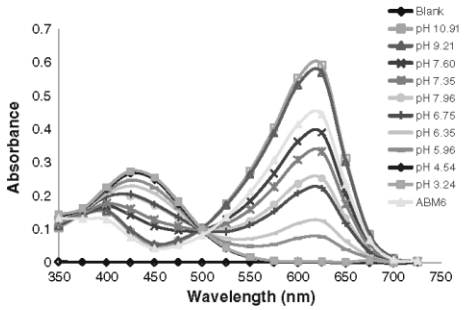


Fig. 3 Absorbance spectrum of bromothymol blue at varying pH values

change over the pH range of 7.6 (blue) to 6 (yellow) with shades of green in the transition region. Figure 4 shows a plot of illumination against pH for the same pH-adjusted samples used in Fig. 3. These data confirm that we can measure the colour change in samples over the 7.6–6 pH range specified by the supplier.

Once the assay configuration had been established and calibrated, the first trial of the system was performed using air-dried cells (dried from 50 µl of liquid consisting of 25 µl overnight culture plus 50 µl 20% *w/v* lactose; see Materials and methods above) following the addition of 1 ml ABM6 amended with various concentrations of arsenic (as sodium arsenate). Figure 5 shows the colour development over 60 h. The tubes are arranged in groups of three replicates in increasing concentration on the *X* axis and time on the *Y* axis. The general trend in the image shows that the higher concentration samples change from blue to yellow more rapidly than the samples with lower arsenic concentration. Note that one tube in the 10 and the 50 ppb arsenic concentrations did not change colour, indicating that further attention must be paid to quality control during preparation of assays.

In Fig. 5, colour changes can be observed over time. The exact moment of this change can be plotted if a quantitative measure of the colour of the tube can be obtained. We measured the tube colour using the quantitative method previously described in the Materials and methods section. Figure 6 shows the quantified change in colour of the pH samples over time. The data were processed by first normalising the cell-free controls to a pH of 7.6 as these tubes do not change colour and are therefore suitable for use as a datum and for colour correction. The tubes are adjusted to the same starting pH value, which corresponds to a pH of 7.6 or above. As the experiment is conducted

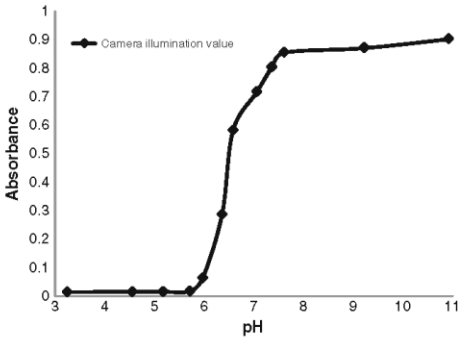


Fig. 4 Numerical illumination value against pH for a range of pH-adjusted samples. Each data point represents a software processed illumination value for a pH calibrated 1 ml tube of water and bromothymol blue

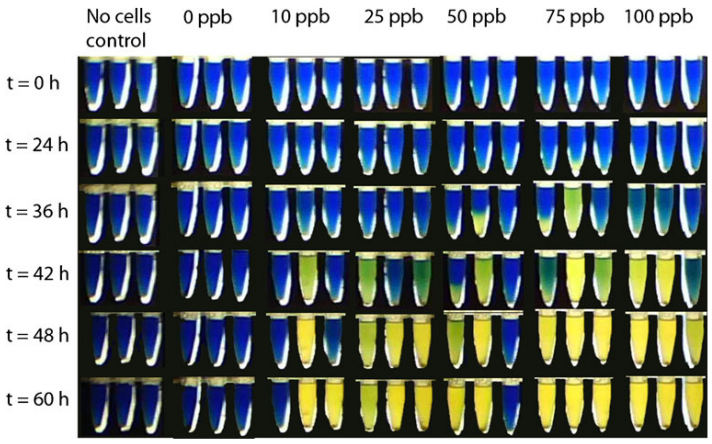


Fig. 5 Arsenic assays using bromothymol blue as a pH indicator. Each arsenic concentration was tested in triplicate on two separate days. Samples (50 µl inoculum) were statically incubated at 37 °C for a period of 65 h. *Blue* represents a safe level of arsenic while *yellow* indicates contamination. This figure shows the biosensor tubes in increasing arsenic concentrations at t=0, 24, 36, 42, 48 and 60 h where the error represents the standard error of six replicates

over a period of several days, these tubes are also used for illumination correction as minor intensity variations are observed due to the day night cycle.

The data shown are the average of five or six replicates, as any tubes that failed to grow were removed for the purpose of data processing. The 50-ppb samples begin to change colour first followed by the 75 and 100 ppb samples. Finally, the 10 and 25 ppb samples change colour at approximately 30 h, about 10 h after the other samples. The arsenic-free control does not exhibit any colour change during the course of the experiment, as also seen in Fig. 5. Thus the time of colour change gives an approximate indication of the relative arsenic concentration.

Figure 7 shows the relative pH change over time for freeze-dried and air-dried cells. For freeze-dried cells, the pH change begins after 8 h in the samples containing 50 and 100 ppb arsenic and is complete after approximately 15 h. The freeze-dried arsenic-free control begins to show a colour change after 12 h, but does not complete this change until nearly 40 h. For the air-dried cells, assays with 50 and 100 ppb arsenic begin to show a change in colour after approximately 12 h, but do not complete the transition until nearly 35 h. The air-dried arsenic-free control begins to change colour after 18 h but does not finish this colour transition until nearly 60 h. This experiment shows an additional cell-free control with 100 ppb arsenic that showed no pH change during the period of the test. If the duration of the test were limited to 24 h, the freeze-dried tubes would have indicated a positive result at both the 50 and 100 ppb arsenic concentrations while showing almost no change in the absence of arsenic.

In previous experiments, a change in pH has been observed in absence of arsenic. A false-positive result is not acceptable in a field test. With the aim of eliminating colour change in arsenic free samples, we performed the same

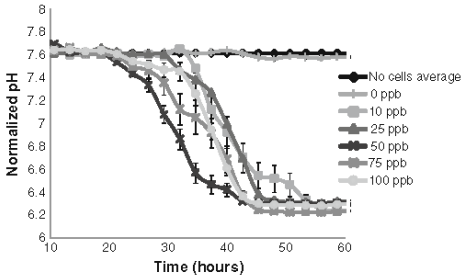


Fig. 6 pH change over time for air-dried cells with varying concentrations of arsenate using bromothymol blue as the indicator. The *error bars* show the standard error calculated from six replicates, where the experiments conducted using three replicates were performed on two separate days

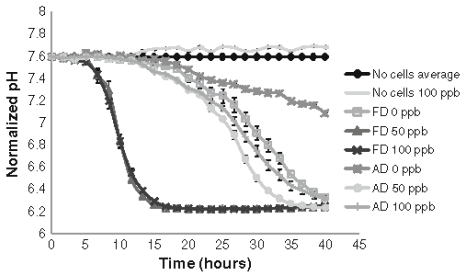


Fig. 7 pH change over time comparing freeze-dried (FD) with air-dried (AD) samples in ABM6 using bromothymol blue as the indicator. Error is the standard error of six replicates

assays using an alternative indicator, bromocresol purple, which has a lower transition pH. Assays with 50 and 100 ppb arsenic responded within 14 and 10 h, while the negative control shows little change until 30 h (Fig. 8). The pH change in the arsenic-free control also appears to halt at around 50 h, before the colour change is complete, suggesting that further modifications to the medium and indicator formulation may yield an assay which does not give a complete colour change in the absence of arsenic even after prolonged incubation.

Effect of other groundwater components

The British Geological Survey [19, 22] report on arsenic contamination in the example region of Bangladesh examined groundwater composition including several ions which could potentially affect the sensor. The *arsR* promoter/operon used in our system is known to be affected by antimonite [23]. As antimonite is also toxic to humans, detecting this compound would not necessarily be considered a false-positive. However, 90% of the samples tested in Bangladesh for antimonite by the British Geological Survey had 0.1 ppb or less, far below the WHO standard and far below the concentration of arsenic [19]. Therefore, antimonite is not thought to be of concern.

Iron (II) and zinc were also considered as possible sources of interference. The BGS report listed the maximum concentrations as 60 and 0.20 mg l^{-1}. Variants of ABM6 media were made with 1,500 mg l^{-1} iron (II) sulphate (527 mg l^{-1} iron) and 2 mg l^{-1} zinc sulphate (0.76 mg l^{-1} zinc) concentrations, respectively, representing nine times and four times the highest reported concentrations. These ions were tested with three air-dried cell replicates on two separate days, and the data are shown in Fig. 9. The data show that there are no noticeable changes in the sensor performance in the presence of zinc and iron (II).

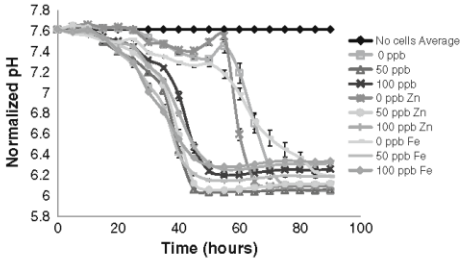


Fig. 9 Effects of iron (II) and zinc on sensor performance using air-dried cells. Samples contained 0, 50 or 100 ppb arsenic as sodium arsenate, in the presence or absence of 0.76 ppm zinc or 527 ppm iron (II) as indicated. Error shown is the standard error of the six replicates

Sensor performance using contaminated groundwater samples from Hungary

To test the biosensor system in a real world situation, we acquired several water samples from wells in Hungary that are known to be contaminated with arsenic [24]. We tested three samples of known concentrations, 1, 48 and 98 ppb arsenic. ABM6 medium at twice the normal strength was diluted 1:1 with filter sterilised water samples to give final arsenic concentrations of 0.5, 24 and 49 ppb arsenic and added to tubes containing air-dried cells. Figure 10 shows results obtained using the high-throughput colorimetric assay. The 24 and 48 ppb samples begin to transition at about 24 h, while the 0.5 ppb sample starts to change colour after 48 h. The arsenic-free control did not significantly change colour during the experiment.

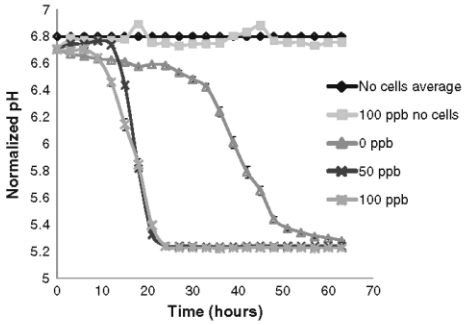


Fig. 8 Freeze-dried cells grown in ABM6 over time in absence and presence of arsenic using bromocresol purple as the indicator. Error is the standard error of three replicates

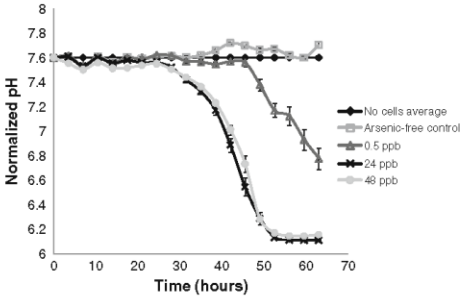


Fig. 10 Responses to arsenic contaminated groundwater samples from Hungary using air-dried cells. Effective concentrations of arsenic were 0.5, 24 and 48 ppb. *Error bars* represent the standard error of the six samples

Discussion

There is a clear need for a cheap, simple, sensitive and reliable assay for monitoring of arsenic concentrations in groundwater in rural areas. Ideally, this should be easy to manufacture, should not require expensive laboratory equipment and should allow a non-expert user to obtain reliable results. A well-designed whole-cell biosensor system could fulfil these requirements. Other groups have described whole-cell biosensors using GFP- and carotenoid-based systems. Tani et al. [25] have described a system using recombinant *E. coli* transformed with plasmids with copies of the *ars* promoter/operator using GFP as a reporter. Although they reported some success obtaining a detection limit of 7.5 ppb, when translated into a field device, a UV excitation and visualisation system is necessary to read the result. Yoshida et al. have also developed a whole-cell arsenite biosensor using *Rhodopseudomonas palustris* with the novel function of having a result that is readable using the human eye [26]. They successfully demonstrated the evaluation of bacterial colour change in the 1.0 to 500 ppb range using 30 volunteers during a double-blind test. Our pH-based biosensor system has the advantage that it gives a clear visual response instead of a relative colour change and shows high sensitivity, responding reliably to arsenate concentrations below the recommended WHO limit of 10 ppb arsenic. We have shown that buffer ions likely to be present in relevant groundwater do not interfere with the assay results and have demonstrated that our system functions with contaminated groundwater samples from Hungary.

One potential issue with a biosensor of this type is that contamination with other lactose-degrading organisms, such as coliforms, may lead to false-positive responses. If the assay is to be used in regional laboratories, this can easily be resolved by autoclaving or filter sterilisation of the samples. For field use, we are designing a disposable device incorporating a sterile filter, so that assay medium and water sample are filter sterilised as they are added to the dried cell pellet. The device can then simply be left in a warm place overnight to allow colour development.

Clearly, before any such device is put into large-scale use, it will be necessary to thoroughly investigate the effects of various parameters such as fluctuations in incubation temperature and presence of other dissolved substances. Our new multi-tube visual monitoring system will allow us to conduct large numbers of assays simultaneously, greatly facilitating our characterization experiments. Since the hardware required is relatively inexpensive, such a system may also be useful for performing large numbers of simultaneous routine assays in regional laboratories. In the current configuration, it appears that the time at which colour change occurs is the most useful parameter for estimating arsenic concentration and that these changes take place within 24 h when freeze-dried cells are used. Our system allows this information to be extracted by processing of the image data. Further development of the imaging protocols could be conducted in an open-source environment and potentially implemented as a plug-in for ImageJ or as a smart phone application.

In terms of storage and distribution, we have demonstrated that the assay works well with cells freeze-dried from lactose solution, and even air-dried cells are usable for at least several weeks following storage at room temperature. We are currently undertaking tests of stability under different storage conditions and have demonstrated successful detection of arsenic in groundwater samples from Hungary. Ultimately, we hope that an assay of this type will allow cheap and simple routine monitoring of arsenic concentrations in groundwater, reducing the burden of illness associated with this widespread problem.

Acknowledgments The authors would like to thank Jovana Radulovic for discussions and advice regarding the analysis of the numerical data, William Edwards for Matlab programming advice, David Kelly for help with Image Pro 7.0 and Steve Mowbray for assistance with lyophilisation. We would also like to thank the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council for funding.

References

1. Guha Mazumder DN, Haque R, Ghosh N, De BK, Santra A, Chakraborty D, Smith AH (1998) Arsenic levels in drinking water and the prevalence of skin lesions in West Bengal, India. Int J Epidemiol 27(5):871–877
2. Fendorf S, Michael HA, van Geen A (2010) Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. Science 328(5982):1123–1127
3. Argos M, Kalra T, Rathouz PJ, Chen Y, Pierce B, Parvez F, Islam T, Ahmed A, Rakibuz-Zaman M, Hasan R, Sarwar G, Slavkovich V, van Geen A, Graziano J, Ahsan H (2010) Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. Lancet 376 (9737):252–258
4. Bhagure GR, Mirgane SR (2010) Heavy metal concentrations in groundwaters and soils of Thane Region of Maharashtra, India. Environ Monit Assess
5. ATSDR (2007) Toxicological Profile for Arsenic. US Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry, Atlanta
6. Hassan MM, Atkins PJ, Dunn CE (2005) Social implications of arsenic poisoning in Bangladesh. Soc Sci Med 61(10):2201–2211
7. Aggett J, Aspell AC (1976) The determination of arsenic (III) and total arsenic by atomic-absorption spectroscopy. Analyst 101 (1202):341–347
8. Sankararamakrishnan N, Chauhan D, Nickson RT, Tripathi RM, Iyengar L (2008) Evaluation of two commercial field test kits used for screening of groundwater for arsenic in Northern India. Sci Total Environ 401(1–3):162–167

9. Kinniburgh DG, Kosmus W (2002) Arsenic contamination in groundwater: some analytical considerations. Talanta 58(1):165–180
10. Sanllorente-Méndez S, Domínguez-Renedo O, Arcos-Martínez J (2010) Immobilization of acetylcholinesterase on screen-printed electrodes. Application to the determination of arsenic(III). Sensors 10(3):2119–2128
11. Sarkar P, Banerjee S, Bhattacharyay D, Turner AP (2010) Electrochemical sensing systems for arsenate estimation by oxidation of L-cysteine. Ecotoxicol Environ Saf 73(6):1495–1501
12. Daunert S, Barrett G, Feliciano JS, Shetty RS, Shrestha S, Smith-Spencer W (2000) Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes. Chem Rev 100(7):2705–2738
13. Stocker J, Balluch D, Gsell M, Harms H, Feliciano J, Daunert S, Malik KA, van der Meer JR (2003) Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water. Environ Sci Technol 37(20):4743–4750
14. Harms H, Rime J, Leupin O, Hug SJ, van der Meer JR (2005) Effect of groundwater composition on arsenic detection by bacterial biosensors. Microchim Acta 151:217–222
15. Diesel E, Schreiber M, van der Meer JR (2009) Development of bacteria-based bioassays for arsenic detection in natural waters. Anal Bioanal Chem 394(3):687–693
16. Trang PT, Berg M, Viet PH, Van Mui N, Van Der Meer JR (2005) Bacterial bioassay for rapid and accurate analysis of arsenic in highly variable groundwater samples. Environ Sci Technol 39 (19):7625–7630
17. Aleksic J, Bizzari F, Cai Y, Davidson B, de Mora K, Ivakhno S, Seshasayee SL, Nicholson J, Wilson J, Elfick A, French CE, Kozma-Bognar L, Ma H, Millar A (2007) Development of a novel biosensor for the detection of arsenic in drinking water. IET Synth Biol 1(1-2):87–90
18. Joshi N, Wang X, Montgomery L, Elfick A, French CE (2009) Novel approaches to biosensors for detection of arsenic in drinking water. Desalination 248(1–3):517–523
19. Kinniburgh DGE (2001) Arsenic contamination of groundwater in Bangladesh, vol 2: final report. British Geological Survey
20. Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor, New York
21. Cowan ST, Steel KJ, Barrow GI, Feltham RKA (1993) Cowan and Steel's manual for the identification of medical bacteria, 3rd edn. Cambridge University Press, Cambridge
22. Kinniburgh DGE (2001) Arsenic contamination of groundwater in Bangladesh, vol 3: hydrochemical atlas. British Geological Survey
23. Rosenstein R, Peschel A, Wieland B, Gotz F (1992) Expression and regulation of the antimonite, arsenite, and arsenate resistance operon of *Staphylococcus xylosus* plasmid pSX267. J Bacteriol 174(11):3676–3683
24. Varsanyi I, Fodre Z, Bartha A (1991) Arsenic in drinking water and mortality in the Southern Great Plain, Hungary. Environ Geochem Health 13:14–22
25. Tani C, Inoue K, Tani Y, Harun-ur-Rashid M, Azuma N, Ueda S, Yoshida K, Maeda I (2009) Sensitive fluorescent microplate bioassay using recombinant *Escherichia coli* with multiple promoter-reporter units in tandem for detection of arsenic. J Biosci Bioeng 108(5):414–420
26. Yoshida K, Inoue K, Takahashi Y, Ueda S, Isoda K, Yagi K, Maeda I (2008) Novel carotenoid-based biosensor for simple visual detection of arsenite: characterization and preliminary evaluation for environmental application. Appl Environ Microbiol 74(21):6730–6738

Appendix 4: Publication in the Journal of Biological Engineering

Jason R Kelly, Adam J Rubin, Joseph H Davis, Caroline M Ajo-Franklin, John Cumbers, Michael J Czar, Kim de Mora, Aaron L Glieberman, Dileep D Monie and Drew Endy

Journal of Biological Engineering BioMed Central

Methodology **Open Access**

Measuring the activity of BioBrick promoters using an in vivo reference standard

Jason R Kelly*[1], Adam J Rubin[1], Joseph H Davis[2], Caroline M Ajo-Franklin[3], John Cumbers[4], Michael J Czar[5], Kim de Mora[6], Aaron L Glieberman[7], Dileep D Monie[8,9] and Drew Endy[1,10]

Address: [1]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 68-580, Cambridge Massachusetts 02139, USA, [2]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge Massachusetts 02139, USA, [3]Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA, [4]Molecular Biology, Cell Biology, and Biochemistry, Brown University, Box G-W, Providence, Rhode Island, 02912, USA, [5]Virginia Bioinformatics Institute, Virginia Polytechnic and State University, Washington Street, MC0477, Blacksburg, VA 24061, USA, [6]Institute for Materials and Processes, The University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, Scotland EH9 3JL, UK, [7]Biomedical Engineering, Division of Engineering, Brown University, Providence, Rhode Island 02912, USA, [8]Department of Ophthalmology, Johns Hopkins University School of Medicine, 600 N Wolfe St, Baltimore, Maryland 21287, USA, [9]Profectus Biosciences, Inc., 6411 Beckley St, Baltimore, Maryland 21224, USA and [10]Department of Bioengineering, Stanford University, Stanford California 94305 USA

Email: Jason R Kelly* - jrkelly@gmail.com; Adam J Rubin - adamjr@mit.edu; Joseph H Davis - jhdavis@mit.edu; Caroline M Ajo-Franklin - CAjo-Franklin@lbl.gov; John Cumbers - johncumbers@gmail.com; Michael J Czar - mczar@vbi.vt.edu; Kim de Mora - kim.demora@gmail.com; Aaron L Glieberman - aaron.glieberman@gmail.com; Dileep D Monie - dmonie@gmail.com; Drew Endy - endy@mit.edu

* Corresponding author

Published: 20 March 2009

Journal of Biological Engineering 2009, 3:4 doi:10.1186/1754-1611-3-4

Received: 4 May 2008
Accepted: 20 March 2009

This article is available from: http://www.jbioleng.org/content/3/1/4

Abstract

Background: The engineering of many-component, synthetic biological systems is being made easier by the development of collections of reusable, standard biological parts. However, the complexity of biology makes it difficult to predict the extent to which such efforts will succeed. As a first practical example, the Registry of Standard Biological Parts started at MIT now maintains and distributes thousands of BioBrick™ standard biological parts. However, BioBrick parts are only standardized in terms of how individual parts are physically assembled into multi-component systems, and most parts remain uncharacterized. Standardized tools, techniques, and units of measurement are needed to facilitate the characterization and reuse of parts by independent researchers across many laboratories.

Results: We found that the absolute activity of BioBrick promoters varies across experimental conditions and measurement instruments. We choose one promoter (BBa_J23101) to serve as an in vivo reference standard for promoter activity. We demonstrated that, by measuring the activity of promoters relative to BBa_J23101, we could reduce variation in reported promoter activity due to differences in test conditions and measurement instruments by ~50%. We defined a Relative Promoter Unit (RPU) in order to report promoter characterization data in compatible units and developed a measurement kit so that researchers might more easily adopt RPU as a standard unit for reporting promoter activity. We distributed a set of test promoters to multiple labs and found good agreement in the reported relative activities of promoters so measured. We also

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

characterized the relative activities of a reference collection of BioBrick promoters in order to further support adoption of RPU-based measurement standards.

Conclusion: Relative activity measurements based on an in vivoreference standard enables improved measurement of promoter activity given variation in measurement conditions and instruments. These improvements are sufficient to begin to support the measurement of promoter activities across many laboratories. Additional in vivo reference standards for other types of biological functions would seem likely to have similar utility, and could thus improve research on the design, production, and reuse of standard biological parts.

Background

The engineering of many-component, synthetic biological systems is being made easier by the development of collections of reusable, standard biological parts [1-6]. Standardization of components has been instrumental in managing complexity in other engineering fields by helping engineers to reliably design and deploy systems comprised of combinations of parts [7]. However, it is an open question whether the overwhelming complexity of living systems will prevent biological engineers from fully achieving similar capabilities (below). To help answer this question, a Registry of Standard Biological Parts started at MIT now maintains and distributes thousands of BioBrick standard biological parts [8]. BioBrick parts provide the first popular example of standard biological parts. However, BioBrick parts are currently only standardized in terms of how individual parts are assembled into multi-component systems (that is, "physical composition") [1,9].

The utility of so-called standard biological parts would increase if the behavior of parts, both in isolation and in combination, were more predictable (that is, "functional composition")[9]. Prediction of behavior, in turn, depends on the initial designs and refinement of the parts themselves, the characterization of part functions, and the representation of part functions via abstract models (for related examples see [10-12]). Today, most BioBrick parts are directly derived from natural DNA sequences with only slight modifications to support at least one physical assembly standard, and many parts remain to be characterized. For example, fewer than 50 out of over 500 transcriptional promoters now available via the Registry have been characterized. Making matters worse, for the 50 characterized promoters, the methods of characterization are disparate and the resulting data incomparable. Shared and standardized approaches are needed in order to begin to address the challenge of characterizing promoters (and other types of standard biological parts) across a distributed community of biological engineers.

Making reliable and comparable in vivomeasurements of biological parts has proven challenging. For example, five different efforts to measure the abundances of proteins in the yeast pheromone mating response system, one of the best characterized eukaryotic signalling systems, produced reports for the numbers per cell (abundances) of key system proteins that vary over a factor of ~12 [Thomson TM, Benjamin KR, Bush A, Love T, Pincus D, Resnekov O, Yu R, Gordon A, Colman-Lerner A, Endy D, Brent R: Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range. Molecular Systems Biology. *Unpublished*]. Such examples suggest that measurement of the state or activity of biological systems, whether natural or engineered, may be unlike past engineering experiences, in that the minor differences in experimental conditions (relative to what can be readily controlled for, below) may cause large changes in the properties being measured. Even if conditions could be controlled for, it has proven challenging for researchers to develop and adopt standard approaches for characterizing biological parts. For example, an analysis of 80 published papers in which researchers used beta-galactosidase (β-gal) activity as a measure of gene expression found that at least six different protocols were used to measure enzyme activity [13]. In addition, nearly all activities were reported in "Miller units" even though in several cases there were differences in the substrates used to quantify enzymatic activity (CPRG or ONPG), the experimental conditions (pH and temperature for the assay), and even the absolute units of the Miller unit (nmol/min or μmol/min) [14]. Differences in conditions such as using either CPRG or ONPG as a substrate for enzymatic assays lead to incompatible results [15], and thus Miller units should generally not be considered comparable unless they have been calibrated against a common reference standard [13].

The challenge of making reliable in vivomeasurements of biological parts is further compounded by the need to measure many part properties indirectly via biological "measurement instruments" such as reporter proteins whose production can also be sensitive to experimental conditions. For example, β-gal activity can be used as an indirect measure of the behavior of a promoter, but the translation and activity of the β-gal protein is itself sensitive to experimental conditions such as temperature or choice of media. Since both the measurement instrument

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

(β-gal) and the property being measured (promoter activity) are sensitive to measurement conditions (perhaps in differing ways) correcting for errors in measured promoter activity due to changes in conditions is more difficult. In theory such challenges could be addressed by strict adherence to standard measurement conditions. However, the adoption of standard measurement conditions in biological engineering is prevented by both practical constraints (as noted above) and also engineering constraints, such as culture or performance requirements that are specific to a particular biotechnology application. The overall situation is summed up nicely via the following quote: "There is no such thing as a standard (biological) component, because even a standard component works differently depending on the environment" [16].

Although the characterization of standard biological parts is challenging, lessons from the measurement of other types of physical objects are worth considering. For example, one approach to controlling for variation in the measured property of an object in response to changing experimental conditions is to collect data from which to develop a model that describes the relationship between the measured property and experimental conditions. As a specific example, models based on empirically determined coefficients of thermal expansion for common building materials (for example, Oak = 54E-6/K at 20°C; Stainless Steel = 17.3E-6/K at 20°C) are now sufficient to enable the reliable construction of structures across a range of environments [17]. However, given the complexity of living matter, the relationships between the measured properties of biological parts and experimental conditions may be difficult to determine (at first). Thus, a second lesson worth considering is the measurement of relative (or ratio) properties rather than absolute characteristics. A relative measure is the ratio of the measurement of some aspect of the object being characterized in comparison to a standard reference object that is measured under the same conditions. For example, early methods for the diagnosis of osteoporosis made use of a measure of spinal cord deformity that was based on the ratio of various length measurements of vertebra within an individual patient [18]. Doctors, by using a relative measurement for length, could account for variation in vertebra sizes between individuals of different body types or heights. As a second example, microarray experiments are frequently performed by co-hybridizing probes synthesized from both a reference and experimental RNA sample that have been labelled with different colors [19]; gene expression levels are then reported as the ratio of the experimental and reference intensities on each array spot. Thus, measurements made in relation to defined reference standards may provide an important first approach in characterizing the in vivoactivity of biological parts and, over time, could enable the collection of empirical data sufficient to support the development of models that describe the effect of varying conditions on part properties.

Here, we characterized the in vivoactivity of BioBrick promoters in order to evaluate if measuring relative activities might provide a useful initial framework for measuring the activity of standard biological parts across varying conditions. We chose to characterize promoters as a first example since they are ubiquitous in engineered biological systems, relatively well-understood, practically useful to biological engineers, and poorly characterized in the existing BioBrick collection [20,21]. We developed a system that allows indirect measurement of the activity of promoters via observation of the synthesis rate of Green Fluorescent Protein (GFP) encoded by mRNA transcribed from each promoter. Our system requires the use of a quantitative model that allows promoter activity to be estimated from observed rates of GFP synthesis (below). Using this approach we demonstrate that normalizing the apparent absolute activity of a promoter to a defined reference standard promoter can help account for variation in conditions that would otherwise lead to significant differences in reported measurements.

Results

Definitions and models for absolute promoter activity

Our first step in characterizing standard biological promoters was to choose the property or properties whose measure would best support the reuse of such parts by biological engineers. Since the primary use of promoters is to initiate transcription, we chose the rate of transcription initiation as the property to be measured. We next chose the promoter clearance rate as the specific property that best describes transcription initiation; we refer to this property as "promoter activity" throughout. In turn, we defined promoter activity as the number of RNA polymerase molecules that pass by (or clear) the final base pair of the promoter and continue along DNA as an elongation complex. We report promoter activity using the generic unit of "Polymerases Per Second," or PoPS, in place of the more traditional "promoter clearance rate" because reporting activity in PoPS allows promoters to be directly compared to other genetic parts whose functioning impacts elongating polymerases, such as transcription terminators [22]. Other properties of promoters such as the binding constant of RNA polymerase to the DNA encoding the promoter, or the secondary structure of the DNA were not considered; while such properties may be relevant to researchers who are studying or engineering new promoters, our focus here was to support researchers who are characterizing or reusing existing promoters.

Directly measuring PoPS in vivois challenging and, to our knowledge, has not yet been reported. However, by plac-

Journal of Biological Engineering 2009, 3:4 http://www.jbioleng.org/content/3/1/4

ing a promoter upstream of the coding sequence for green fluorescent protein (GFP) we could use the rate of GFP synthesis as an indirect measure of promoter activity. We could then use a quantitative model to relate observed GFP synthesis rates to promoter activities reported as PoPS.

We adopted a previously described ordinary differential equation (ODE) model of GFP expression from a constitutive promoter to relate GFP synthesis rates per cell to promoter activities [9,23]. We evaluated this ODE model at steady-state (Additional file 1) in order to determine the rates of successful mRNA initiation events per DNA copy of each promoter ($PoPS^{SS}$) given observed GFP synthesis rates per cell (S^{SS}_{cell}):

$$Absolute\ activity\ of\ promoter\ (PoPS^{SS}) = \frac{\gamma_M(a+\gamma_I)S^{SS}_{cell}}{\rho a n} \quad (1)$$

where γ_M is the mRNA degradation rate, a is the GFP maturation rate, γ_I is the degradation rate of immature GFP, ρ is the translation rate of immature GFP from mRNA, and n is the number of copies of the promoter in the cell.

Variability due to equipment and conditions

We explored the sensitivity of our observable measure of promoter activity, GFP synthesis rates, to different measurement conditions and different measurement instruments. We estimated the per cell GFP synthesis rates of two promoters (BBa_J23101 and BBa_J23150) across seven different measurement conditions and instruments (Figure 1A). We estimated GFP synthesis rate by reporting the change in arbitrary fluorescence units per absorbance over a 1-hour period in log phase growth (Methods). We varied the experimental conditions by changing the cell strain (TOP10 or W3110), carbon source (glucose or glycerol), and temperature (30C or 37C) during growth. We found that the observed GFP synthesis rates were sensitive to the choice of strain (varying up to 2-fold) but insensitive to temperature and carbon source (within experimental error). We also varied the plasmid copy number and plasmid antibiotic resistance marker in order to explore how different genetic "measurement instruments" might impact the measured GFP synthesis rates. We found that GFP synthesis rates were sensitive to the plasmid copy number (varying up to 3-fold) and antibiotic resistance marker (varying up to 1.5-fold).

Definition and models for relative promoter activities

We noted that the activity of promoters (for example, J23101 and J23150) measured across different conditions or with different instruments was correlated (Figure 1A). This correlation suggested that a measure of relative promoter activity might be less sensitive to varying experimental conditions or measurement instruments. To test this idea, we defined a new property – relative promoter activity – as the ratio of the absolute activity of a sample promoter, ϕ, relative to the absolute activity of a standard reference promoter, BBa_J23101, with both promoters measured under equivalent conditions and with the same measurement instrument. We reported relative promoter activities in a newly defined unit: Relative Promoter Units or RPUs. By our definition, a sample promoter with a relative activity of 1 RPU has activity equivalent to BBa_J23101.

An important consequence of considering a relative unit of measurement for reporting promoter activities is that many of the difficult-to-measure model parameters (Equation 1) that might change with changing environmental conditions can be cancelled when calculating relative promoter activities:

$$Relative\ activity\ of\ promoter\ \varphi\ (RPUs) = \frac{PoPS^{SS}_{\varphi}}{PoPS^{SS}_{J23101}} \quad (2)$$

Thus, by substituting Eq. 1 into Eq. 2, we calculated promoter strengths in relative units of RPU and thereby eliminated many of the elementary parameters found in Eq. 1 via a cancellation of terms:

$$Rel.\ activity\ of\ promoter\ \phi\ (RPU) = \frac{\frac{\gamma_{M,\phi}(a_\phi+\gamma_{I,\phi})S^{SS}_{cell,\phi}}{\rho_\phi a_\phi n_\phi}}{\frac{\gamma_{M,J23101}(a_{J23101}+\gamma_{I,J23101})S^{SS}_{cell,J23101}}{\rho_{J23101}a_{J23101}n_{J23101}}} \quad (3)$$

We then made four additional assumptions that further simplified Eq. 3. First, we assumed that GFP expressed from either the test promoter ϕ or the reference standard promoter has an equivalent maturation rate ($a_\phi = a_{J23101} = a$; given that the two promoters are measured under the same culture conditions). Second, since both promoters are carried on the same backbone plasmid, we assumed that each promoter is at the same average copy number ($n_\phi = n_{J23101}$); while there are reported cases of promoter activity influencing the copy number of plasmids due to RNA polymerases transcribing through the plasmid origin of replication [24], a transcription terminator (BBa_B0015) downstream of our test construct's GFP coding sequence as well as the transcription terminators flanking the BioBrick cloning site [25] should largely prevent differences in promoter activity from impacting plasmid copy number. Third, since the promoters tested here have been standardized to have identical transcription initiation sites (predicted) and identical sequences down-

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

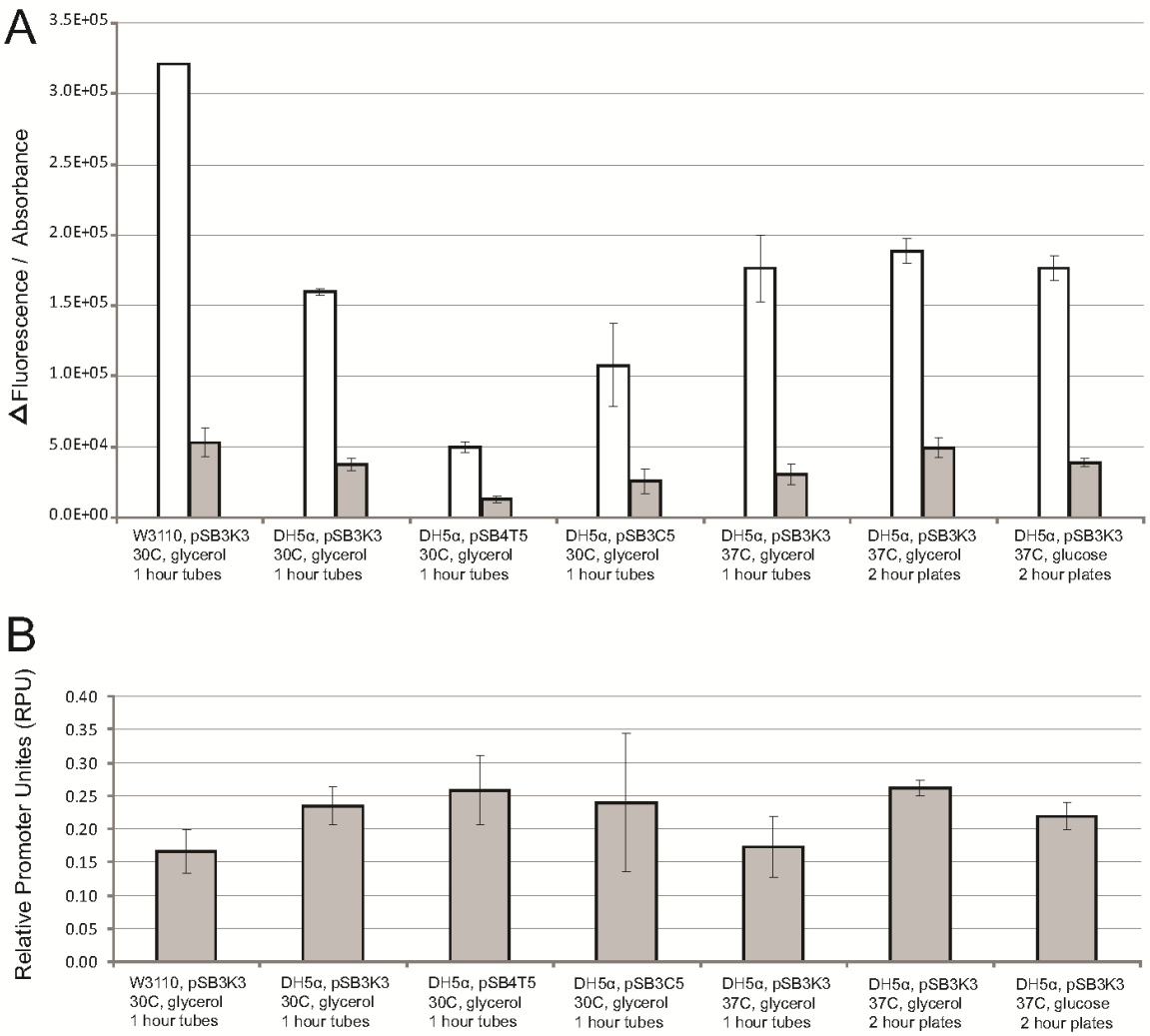


Figure 1

Reference standards reduce variation in reported promoter activities under different measurement conditions. We measured the activity of 2 promoters, J23101 (white columns) and J23150 (grey columns) under seven different measurement conditions and measurement instruments. We varied the media, temperature, cell strain, and plasmid copy number of the promoter test construct. (A) To estimate the per cell GFP synthesis rate we reported the change in fluorescence over a 1 hour period in exponential phase divided by the average absorbance during this period. The coefficient of variation of the GFP synthesis rates across the seven measurement approaches was 49% for J23101 and 39% for J23150. (B) We used the same data and divided the GFP synthesis rate of J23150 (grey bars) by that of J23101 (white bars) in order to calculate the relative promoter activity of J23150 in RPUs. The coefficient of variation of the relative promoter activity across the seven measurement approaches was only 17% suggesting that the relative promoter activity is less sensitive to conditions then absolute activity measured by per cell GFP synthesis rate.

stream of the initiation site (Additional file 1) we expected that each promoter produces the same mRNA sequence [26]. Since the transcribed mRNAs are expected to be identical we assumed that their mRNA degradation rates are equivalent ($\gamma_{M,\phi} = \gamma_{M,J23101}$) and that the translation rates of immature GFP from mRNA are also equivalent ($\rho_{\phi} = \rho_{J23101}$); while mRNA degradation is also a function of dilution due to cell growth, the dilution rate is negligible

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

relative to typical rates of active mRNA degradation in *E. coli* [27]. Finally, we assumed that immature GFP is stable so that protein degradation is negligible compared to dilution due to cell growth ($\gamma_{I,\phi} = \mu_\phi$ and $\gamma_{I,J23101} = \mu_{J23101}$, where μ is the cellular growth rate). Following the above assumptions, we simplified Eq. 3 to:

$$Relative\ activity\ of\ promoter\ \phi\ (RPU) = \frac{(a+\mu_\phi)S^{SS}_{cell,\phi}}{(a+\mu_{J23101})S^{SS}_{cell,J23101}} \quad (4)$$

We further simplified Eq.4 by noting that:

$$if\ |\mu_\varphi - \mu_{J23101}| \ll a\ then\ \frac{(a+\mu_\varphi)}{(a+\mu_{J23101})} \approx 1 \quad (5)$$

For example, we measured the growth rates of cells in our experiments to determine if the difference between the growth rates of cells containing the promoter test construct (μ_φ) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the maturation rate of GFP (that is, $|\mu_\varphi - \mu_{J23101}| << a$). The cellular growth rates varied depending on the promoter being tested as well as on the experimental conditions: the fastest growth rate was observed in cells containing the BBa_J23113 promoter test construct grown in M9+glucose (μ = 0.9 hr^{-1}), and the slowest growth rate was observed in cells containing the BBa_R0040 promoter test construct grown in M9+glycerol (μ = 0.5 hr^{-1}). The maturation rate of the GFP variant used in the GFP reporter device (BBa_E0040) has been measured previously as a = 6.48 hr^{-1}[9]. Based on the worst-case assumption that cells containing the promoter test construct are the fastest growing cells (μ_φ = 0.9 hr^{-1}), and that cells containing the reference standard construct are the slowest growing cells (μ_{J23101} = 0.5 hr^{-1}) then:

$$\frac{(a+\mu_\phi)}{(a+\mu_{J23101})} = 1.06 \approx 1 \quad (6)$$

Therefore, we assumed that the difference between the growth rates of cells containing the promoter test construct (μ_φ) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the maturation rate of GFP, allowing Eq. 4 to be combined with Eq. 5 yielding:

$$Relative\ activity\ of\ promoter\ \phi\ (RPU) = \frac{S^{SS}_{cell,\phi}}{S^{SS}_{cell,J23101}} \quad (7)$$

Taken together, by reporting promoter activity relative to a reference standard promoter (BBa_J23101) and choosing promoters with identical transcription initiation sites and identical sequences downstream of the initiation sites, researchers can quickly report measured relative promoter activities in compatible units without having to independently measure GFP maturation rates, mRNA degradation rates, protein production rates, or plasmid copy number for their specific experimental setup. (We detail the precise numerical sensitivity of the quantitative model to each of the above assumptions in Additional file 1).

We converted GFP synthesis rates measured across 7 different conditions and instruments (Figure 1A; Coefficient of variation (CV) of the measurements is 39.1%) to relative promoter activity in RPUs (Figure 1B; CV of the measurements is 17.5%). We noted that the coefficient of variation in promoter activity was reduced by approximately half when converted to RPUs from GFP synthesis rates. This reduction in variation suggests that relative promoter activity might be a useful property for characterizing promoters. However, care should be taken to note that while relative promoter activities remain fairly constant across some range of conditions, absolute promoter activities vary widely across these same conditions. Stated differently, a promoter that has an equivalent relative activity across multiple conditions might not produce equal absolute activity (as measured in PoPS) across the same conditions (please see Discussion).

Laboratory-laboratory variation

Our initial success in characterizing relative promoter activity across different conditions and measurement instruments suggested a practical test. Specifically, we sought to determine whether multiple laboratories could work together to characterize promoters. To do this, we distributed a "reference promoter set" comprised of four strains, each containing one promoter test construct (BBa_J23113, BBa_J23150, BBa_J23151, or BBa_J23102) to researchers in six independent laboratories. Each researcher then measured the activity of the four promoters following a five-step procedure: (1) three independent cultures were grown from single colonies for each of the four promoters, (2) cells were collected in exponential phase, (3) GFP concentration per cell was measured using a flow cytometer, (4) flow cytometer data was gated based on forward and side scatter and the negative control, and (5) the geometric mean of the per cell fluorescence in the population was reported for each culture (Methods). We made no efforts to standardize the equipment (flow cytometers) or equipment settings beyond asking researchers to use typical settings for measuring GFP and by providing each lab with an example plot to guide gating of the flow cytometry data based on forward scatter, side scatter, and fluorescence [28]. As expected, there were slight differences in how the protocol was conducted in each laboratory, such as different culture conditions (roll-

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

ers or shakers) and growth time. Since the measurements were reported in common units of RPUs we are able to compare the results of the interlaboratory promoter activity measurements directly (Figure 2). The mean promoter activity measured by each lab is relatively consistent across all laboratories with less than a 2-fold range of activities (min-max) across all measured promoters (BBa_J23150: 0.14–0.23 RPU; BBa_J23150: 0.38–0.61 RPU; BBa_J23103: 0.77–0.96 RPU). The activity of the weakest promoter, BBa_J23113, was equivalent to the negative control within error for all but one of the laboratories. These results suggest that relative promoter activity is an effective metric for making comparable measurements across multiple laboratories. Finally, we determined the coefficients of variation of the measured promoter activities across all labs to be 17.2%, 17.1%, and 8.5% for BBa_J23150, BBa_J23151, and BBa_J23101, respectively, setting a baseline for future improvements to the measurement kit and methods.

Community-based measurement of promoter collections

Given that many laboratories could coordinate their measurement of promoter activities, we sought to prepare tools that would facilitate the widespread adoption of relative promoter activity measurements. To do this we first measured the relative activities of a set of seven representative promoters obtained from the Registry of Standard Biological Parts (Figure 3) [8]. These promoters included members of a constitutive promoter library (BBa_J23100 – BBa_J23119, constructed by JC Anderson) as well as the commonly used Tet repressor (BBa_R0040) and Lac repressor (BBa_R0011) regulated promoters [29]. The regulated promoters were tested in the absence of their cognate repressor proteins. Such libraries of characterized promoters have been shown to be valuable to researchers for tuning biochemical networks to optimize the synthesis of products of interest [20,30]. We measured the relative promoter activities by calculating the steady-state GFP synthesis rates (Methods) and converting these rates to RPUs. Nine independent clones were characterized across three separate experimental runs for each promoter tested. The promoters ranged in activity from 0.026 ± 0.003 to

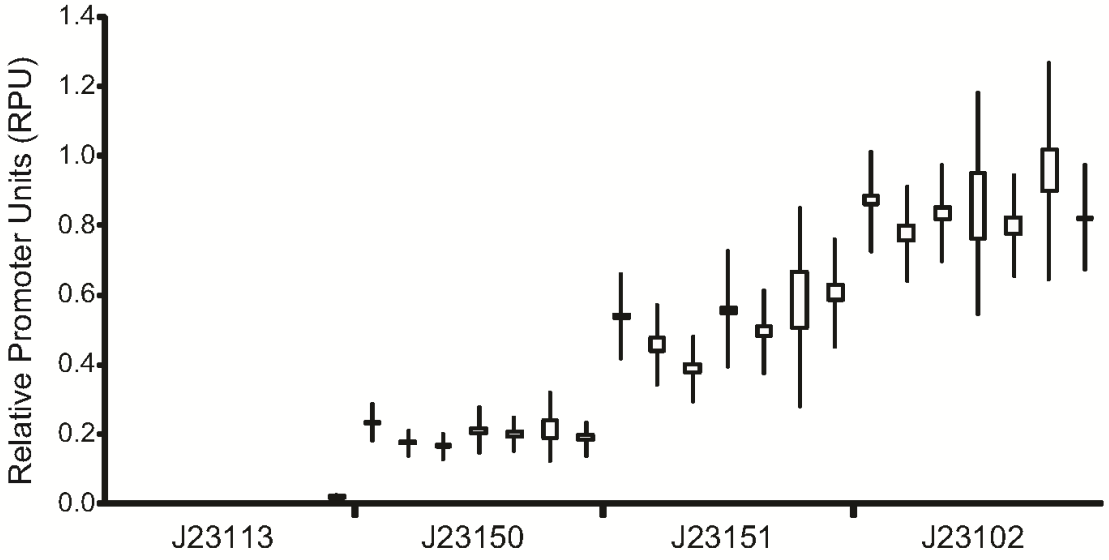


Figure 2

Reference standards and units allow independent labs to make sharable measurements. Each laboratory followed the same measurement procedure, measuring relative promoter activities based on GFP concentration measured via a flow cytometer. Measurements were taken in triplicate; the boxes show the highest and lowest measured relative promoter activities and the whiskers show the 95% confidence interval of the mean of the activities. The large range in the 95% confidence interval (extending beyond the highest and lowest measured activities) is partially a function of the small number of replicates (three) that were conducted by each laboratory. The activity of BBa_J23113 was equivalent to the negative control within error for all but one of the laboratories. The measured activities of the other three promoters were fairly consistent across laboratories with less than a 2-fold range of activities measured for each promoter across all labs (BBa_J23150: 0.14 – 0.23; BBa_J23150: 0.38 – 0.606; BBa_J23103: 0.77 – 0.96).

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

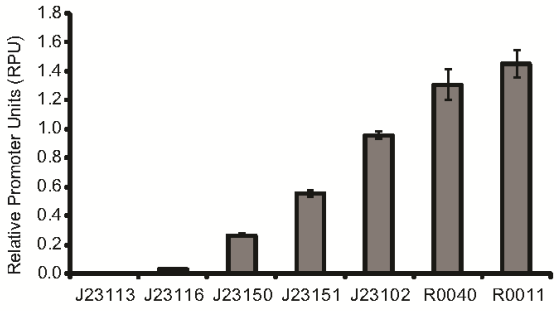


Figure 3
Promoter collections can be readily characterized via Relative Promoter Units (RPUs). The five promoters labeled J23### are from a constitutive promoter library and R0040 and R0011 are tet- and lac-repressible promoters, respectively. The activity of the promoters was measured in relative promoter units (RPUs). This collection of promoter may itself be useful for tuning gene expression in engineered systems. The error bars represent the 95% confidence interval of the mean based on nine replicates.

1.45 ± 0.095 RPUs (uncertainties represent 95% confidence interval of the mean). The GFP expression level from one promoter (BBa_J23113) was statistically equivalent within measurement error to the expression level of the negative control (TOP10).

To further support community-based standardized measurement of promoter activities, we developed a first generation measurement kit for characterizing the relative activity of BioBrick promoters in RPUs. Our overarching objective for the kits was to enable independent researchers to make comparable measurements of relative promoter activity in standard units. We developed instructions and a parts list for the promoter measurement kit (Additional file 1 [Supp Box1 and Supp Table1]). The promoter measurement kit contains measurement "instruments" such as a green fluorescent protein (GFP) reporter device (BBa_E0240) and backbone plasmid (pSB3K3), as well as a recommended *E. coli* strain (TOP10). The reference promoter (BBa_J23101) was inserted upstream of the GFP reporter device (BBa_E0240) and included in the kit as the reference standard construct (BBa_I20260). In order to measure the activity of a user-specified promoter, kit users assemble the user-specified promoter upstream of the GFP reporter device and insert this combined part into the backbone plasmid to form the promoter test construct. The process for inserting a promoter upstream of the GFP reporter device is based on three-antibiotic BioBrick standard assembly [25], and is outlined in the instructions included with the kit (Additional file 1 [Supplementary Box 1]).

Discussion

We found the absolute activity of promoters to vary under different experimental conditions and when using different measurement instruments. We chose a promoter (BBa_J23101) to serve as a reference standard and demonstrated that by measuring relative promoter activity (activity of a sample promoter divided by activity of the reference standard promoter BBa_J23101, measured under the same conditions) we could reduce reported variation in measured promoter activity across differing experimental conditions and equipment. We defined the Relative Promoter Unit (RPU) in order to enable researchers to report promoter characterization results in compatible units, and developed a measurement kit in order to more easily allow researchers to adopt an RPU-based measurement approach. We distributed a test set of 4 promoters to 7 independent labs and found good agreement in the measured relative promoter activities across the test set. Finally, we characterized the relative promoter activity of 7 BioBrick promoters in order to bootstrap a collection of promoters measured according to our initial RPU reference standard.

Absolute and relative promoter activities

The absolute activity of a promoter is defined by the number of elongating polymerases per second (PoPS) exiting the promoter. The same promoter under different environmental conditions can have widely varying absolute promoter activities (Figure 1A). Moreover, it is challenging to relate an indirect measure of absolute promoter activity, such as per cell GFP synthesis rates, to PoPS since any variation across conditions in the functioning of the genetic measurement instrument may not be well correlated with variation in promoter activity across these same conditions. In contrast, the relative activity of a promoter is defined as the ratio of the absolute activity of the promoter to the absolute activity of a reference standard promoter measured under the same conditions and with the same measurement instrument. We found that the relative activity of a promoter will remain fairly constant across a practical range of conditions (Figure 1B).

Relative promoter activity as reported in RPUs enables the ranking of the activities of promoters but does not, by itself, provide information about the absolute activity of the promoter under particular conditions. For example, if the relative activity of a sample promoter remains constant across several conditions, it is not necessarily the case that the promoter is producing equal numbers of mRNA transcripts in each condition, only that the promoter activity is remaining proportional to the reference standard across the conditions. Nonetheless, relative pro-

Journal of Biological Engineering 2009, 3:4 http://www.jbioleng.org/content/3/1/4

moter activity is a valuable property to measure and report since promoters can be rank ordered by relative promoter activity even if they were characterized by different researchers or across different environments (Additional file 1 [Supplementary Table 4]). Furthermore, if both the relative and absolute activities of a promoter are measured, then a conversion factor can be established that defines the relationship between relative and absolute promoter activity under specific measurement conditions. For example, the reference standard promoter J23101, which has a relative activity of 1 RPU, has an estimated absolute activity of ~0.03 PoPS under specific conditions (Additional file 1). Therefore, the promoter J23151, which has a relative activity of approximately 0.5 RPU under these same conditions, would be predicted to have an absolute activity of ~0.015 PoPS. As more absolute measurements are made, an expanded set of conversion factors (or functions) could be developed, allowing for improved estimates of absolute promoter activities across a wider range of measurement conditions.

Interlaboratory measurement of promoter activity

Measurement of in vivopromoter activities across laboratories is challenging due to the sensitivity of results to both experimental conditions and measurement instruments (Figure 1), as well as the lack of shared reference standards [13]. We demonstrated here that by using a relative promoter activity to characterize promoter strength based on a shared reference standard, seven independent laboratories could make comparable measurements of three promoters (Figure 2). We expect that future improvements to the recommended measurement techniques and measurement kit components could further reduce the variation in measurements across laboratories, with the results reported here providing a practical baseline for judging proposed improvements.

Measurement procedures

The reference promoter BBa_J23101 and relative promoter activity measured in RPUs provide a shared platform for researchers to evaluate different measurement procedures. We deliberately have not advocated a single measurement procedure, and there should be many acceptable procedures for characterizing promoters in RPUs, just as units for length or mass are not tied to a single measurement approach. The choice of the best measurement procedure will be influenced by the particular group making the measurements. For example, laboratories without access to equipment for capturing high-throughput single-cell measurements of fluorescence might opt for a bulk fluorescence measurement using a fluorimeter. Other groups might prefer to obtain single-cell measurements using quantitative microscopy or flow cytometry, or to capture a time-course of fluorescence measurements from a growing culture. As different measurement procedures are likely to have merits within different communities we expect that a number of procedures will be established. As an example, for the community of undergraduate teams using the promoter measurement kit during the International Genetically Engineered Machines (iGEM) competition, we have suggested a measurement protocol that can be easily carried out by novice researchers and that only requires two absorbance and bulk fluorescence measurements (Additional file 1 [Supplementary Box 2]) [31].

Engineering with characterized promoters

We anticipate that both absolute and relative promoter measurements will be useful in engineering genetic networks, however it is unclear to what extent one approach might be preferred over the other (presuming both types of measurements could be readily obtained). For example, we can imagine an engineering design framework in which the absolute activities of promoters and other functional genetic elements are tracked explicitly, in order to support detailed modeling and analysis of issues such as the absolute "load" placed on a host cell via recombinant gene expression. Such an ability seems likely to become more important as many-component engineered biological systems are attempted (dozens to hundreds of gene products), in which the absolute expression levels of individual genes must be well managed and might be kept low compared to the high-expression, protein production systems typically used today. However, we can also imagine a competing or complementary engineering framework, based on the idea that cells already provide self-adapting and robust environments within which the absolute activities of genetic elements such as promoters are finely regulated by overall environmental or culture conditions [32]. In such a framework, relative measurements of promoter activities may be both easier to obtain and more relevant. Many natural biological systems already follow this model, and are robust to the absolute properties of components so long as the relative relationships between subparts are maintained. For example, developmental body plans may vary in size with individual organisms having different overall sizes, however the ratio of the sizes of individual bones or organs to overall body mass is often tightly maintained [33].

Standard promoter definition

The promoters tested here were practically standardized to have identical transcription initiation sites (predicted) and identical sequences downstream of the initiation site (Additional file 1). Thus, we expect that the mRNA expressed by each of the tested promoters is identical to the mRNA produced by the reference standard promoter, and that we can cancel the mRNA degradation rate and translation rate of immature GFP from mRNA terms in simplifying the model relating GFP synthesis rates to

Journal of Biological Engineering 2009, 3:4 http://www.jbioleng.org/content/3/1/4

RPUs. This simplification allows the activity of promoters to be reported in comparable units (RPUs) without needing to directly measure mRNA levels. Going forward, an expanded definition for standard BioBrick promoters could be developed in order to ensure that all promoters share the same transcription start position and a fixed sequence downstream of the transcription start site (that is, 5' mRNA UTR). All promoters that adhered to such a standard could then be reliably measured using the kits described here, or via future kits based on gene expression reporters that adhered to any new standard, without the need for promoter-specific mRNA quantitation.

Distribution, use, and improvement of standardized measurement kits

Shared measurement tools and reference standards become more useful as they are broadly adopted. To facilitate such adoption, the Registry of Standard Biological Parts now includes our promoter measurement kit and reference standard in the annual distribution of BioBrick parts. We also created a website in order to support the reporting and sharing of promoter activity measurements http://partsregistry.org/measurement. This website contains instructions for use of the kit and summarizes previously characterized promoters. Finally, to enable discussion of proposed improvements to the kit and reference standard, and also the development of new kits and reference standards, we are supporting an open discussion of technical standards in synthetic biology http://biobricks.org/standards.

Conclusion

Standard tools, techniques, and units for measurement are needed for a distributed community of biological engineers to independently characterize and share biological parts. We have defined a shared unit for measuring relative promoter activity (Reference Promoter Units, RPUs) and demonstrated that relative promoter activity can address some of the challenges in measurement across labs due to varying experimental conditions and measurement instruments. We developed a first-generation measurement kit for BioBrick promoters, and are freely distributing the kit via the Registry of Standard Biological Parts. Having demonstrated the feasibility and ease of use of the kit, we hope to encourage a community of users to adopt and improve these measurement tools and reference standard in order to characterize promoters via a comparable and common unit, the RPU. We expect that the shared experiences of biological engineers using common measurement tools and standards will help to identify new engineering challenges in improving the reliability and reuse of standard biological parts.

Methods

Strains and media

All measurement experiments and cloning were performed in *E. coli* TOP10 (Invitrogen) or W3110. Supplemented M9 minimal medium (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 0.1 M $MgSO_4$, 0.5 M $CaCl_2$) was used for all measurement experiments with either glycerol (0.4%) or glucose (0.4%) added as a carbon source and kanamycin (20 μg/ml) antibiotic added where appropriate. All oligonucleotides were purchased from Invitrogen and DNA modifying enzymes were purchased from New England Biolabs.

Promoter measurement kit contents

Sequences for all BioBrick plasmids (denoted pSB***) and BioBrick parts (denoted BBa_####) are available through the Registry of Standard Biological Parts [8]. pSB3K3 contains a p15A origin of replication (copy number 10–12) and the kanamycin resistance marker, pSB4T5 contains the pSC101 origin of replication (copy number ~5) and the tetracycline resistance marker, and pSB3C5 contains the p15A origin of replication and the chloramphenicol resistance marker. Physical copies of the plasmids and parts are also available from the Registry via the annual Registry parts distribution. The details of the promoter measurement kit contents are described in Supplementary Box 1 and Supplementary Table 1 (see Additional file 1). The sequences for the preparative primers used to amplify pSB3K3 to generate backbone plasmid are: TACTAGTAGCGGCCGCTGCAG (forward primer) and CTCTAGAAGCGGCCGCGAATTC (reverse primer).

Assembly of test constructs

We built promoters by annealing synthesized oligonucleotides. The oligonucleotides were ordered with 5' phosphates and designed to leave an EcoRI overhang on the 5' end and a SpeI overhand on the 3' end so they could be used in subsequent ligation reactions without an intermediate restriction digest step. We inserted seven promoters: BBa_J23113, BBa_J23116, BBa_J23150, BBa_J23151, BBa_J23102, BBa_R0040, and BBa_R0011 into the promoter test construct and transformed into TOP10 according to the process outlined in Supplementary Box 1 (see Additional file 1). We found the optimal concentration of DNA for each of the three components in the ligation reaction (pSB3K3, BBa_E0240 or BBa_I13401, and the test promoter) was approximately 10 ng per uL. More detailed protocols and troubleshooting can be found at http://partsregistry.org/measurement. In the process of construction we found mutations in two of the promoters that we attribute to errors in the synthesis of the oligonucleotides that were annealed to construct the promoters. The two promoters were functional so we included them as additional members of the collection (BBa_J23150 and BBa_J23151). The method of part assembly described

Journal of Biological Engineering 2009, **3**:4 http://www.jbioleng.org/content/3/1/4

here is based on the three-antibiotic BioBrick standard assembly method [25].

Assay of promoter collection

The protocol described here will be referred to as the "original" protocol throughout the methods section and describes the measurement procedure used to characterize the set of seven promoters (Figure 3). For each promoter construct three 17 mm test tubes containing 5 ml of pre-warmed (37°C) supplemented M9 medium with kanamycin (20 μg/ml) were inoculated from single colonies of TOP10-DH5α containing the promoter test construct on the pSB3K3 vector backbone. Cultures were grown in 17 mm test tubes for approximately 20 hrs at 37°C with spinning at 70 rpm. We then diluted the cultures 1:100 into 5 ml of pre-warmed fresh media and the cultures were grown for approximately four hours under the previous conditions (17 mm tubes, 37°C, spinning at 70 rpm). After four hours, we measured the OD600 of a 500 μl aliquot from each culture on a WPA Biowave Spectrophotometer. Based on this OD measurement, the cultures were diluted to the same OD (0.07) in 5 ml of pre-warmed fresh media and grown for one hour at 37°C. We then transferred three 200 μl aliquots from each culture into a flat-bottomed 96 well plate (Cellstar Uclear bottom, Greiner). We incubated the plate in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer) at 37°C and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter, 0.1 second counting time through 5 mm of fluid), fluorescence measurements (485 nm excitation filter, 525 nm emission filter, 0.1 seconds, CW lamp energy 12901 units), and shaking (3 mm, linear, normal speed, 15 seconds).

Background absorbance was determined by measuring wells containing only media. Background fluorescence was determined at different ODs from the fluorescence of TOP10 cells without a GFP expressing vector [34]. After background subtraction, time-series fluorescence (F) and absorbance (ABS) measurements were used to calculate the ratio of the rates of GFP synthesis for the promoter test construct and the reference standard construct. Measurements were taken from an approximately 30 min period in midexponential growth [35] (Additional file 1 [Supplementary Materials Figure 1 & 2]). For example:

$$RPU = \frac{S^{SS}_{cell,\phi}}{S^{SS}_{cell,J23101}} = \frac{(dF_{\phi}/dt)/ABS_{\phi}}{(dF_{J23101}/dt)/ABS_{J23101}} \quad (9)$$

Since we are calculating a ratio of the GFP synthesis rates we do not need to determine each rate in absolute units of GFP per second per cell, rather we can use the background-subtracted fluorescence (F) that is proportional to the number of GFP molecules and the background-subtracted absorbance (ABS) that is proportional to the number of cells in the culture to calculate the ratio of GFP synthesis rates [9,36].

Assay of different measurement conditions

We measured the promoter activity of two promoters (BBa_J23101 and BBa_J23150) under seven different measurement procedures. The first of the seven procedures was identical to the "original" protocol described above for measuring the 7-member promoter collection except it was conducted at 30°C in the strain W3110 with pSB3K3 as the vector backbone for the promoter test construct. The second procedure was identical to the original except it was conducted at 30°C. The third procedure was identical to the original except that it was conducted at 30°C and pSB4T5 was used as the vector backbone. The fourth procedure was identical to the original except that it was conducted at 30°C and used pSB3C5 as the vector backbone. The fifth procedure was identical to the original. The sixth procedure was identical to the original except that instead of the second dilution into tubes followed by 1 hour of growth, the cells were diluted into 96 well plates and incubated for two hours before we started taking measurements. The seventh procedure was identical to the sixth except glucose was used instead of glycerol as the carbon source.

Assay of inter-laboratory variability

We distributed a set of four promoters (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) to six laboratories to take independent measurements of promoter activity. The protocol each lab conducted was identical to the original protocol described, except that the cells were harvested after the first 1:100 dilution and 4 hours of growth (there was no second dilution step). The cells were then spun down, resuspended in PBS, and the fluorescence per cell was measured using a flow cytometer. The measurement equipment used (cytometer model, laser, emission filter) varied between the laboratories (see Additional file 1 [Supplementary Table 2]).

For all other experiments we measured RPUs from the GFP synthesis rates as described in Equation 9. However, for the inter-laboratory experiments we are unable to measure the GFP synthesis rates because these rates require a time series to calculate (dG/dt in Eq. 9) and we only requested a single time point, however we can use this single time point to find the background-subtracted per cell fluorescence at steady-state ([F]). The flow cytometer measures fluorescence per cell directly, thus (F) is calculated by taking the geometric mean of the population fluorescence per cell. We related the per cell GFP concen-

tration ([G]) to RPUs by using a model described previously [23] (derivation in Additional file 1):

$$RPU = \frac{[G]_{cell,\varphi}}{[G]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} = \frac{[F]_{cell,\varphi}}{[F]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} \quad (10)$$

μ_{ϕ}/μ_{J23101} is a correction term based on differences in growth rate. Changes in the growth rate effect per cell GFP accumulation since loss of GFP per cell is largely due to dilution. Since we are calculating a ratio of the per cell GFP concentrations we do not need to determine each rate in absolute units of GFP molecules per cell, rather we can use the background-subtracted per cell fluorescence ([F]) that is proportional to the number of GFP molecules per cell to calculate the ratio of GFP concentrations.

After background correction, the per cell fluorescence ([F]) was determined for each promoter and activities in RPUs were calculated using Eq. 10. We applied the growth rates measured previously (Additional file 1 [Supplementary Table 3]) across all laboratories when calculating RPUs, rather than requesting individual laboratories to measure growth rates. This approximation likely increased the variability in the promoter activity measurements across laboratories, as growth rates vary between laboratories due to differences in culture conditions and media.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JK, AR, and JD designed and conducted all experiments except those done at other laboratories. JK, KD, and CA designed the multi-institution experiments which were conducted by AR, CA, JC, MC, KD, AG, and DM. JK and DE conceived of the study and wrote the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Supplementary material. Various supplemental materials

Click here for file

[http://www.biomedcentral.com/content/supplementary/1754-1611-3-4-S1.doc]

Acknowledgements

The authors would like to thank Nishant Baht and Karen Wong for building and testing alternate designs of the measurement kits. The authors would like to thank Barry Canton for helpful discussions. The authors would like to thank Julius B. Lucks, Lei Qi, Weston Whitaker and Gavin Price for taking measurements in the multi-laboratory study. JC and AG would like to acknowledge: Stephanie Terrizzi and the Flow Cytometry and Cell Sorter Facility for flow cytometry (FACSAria funded by NCCR SIG grant # 1S10RR021051-01A2). CA would like to recognize that the work was performed in part at the Molecular Foundry, Lawrence Berkeley National Laboratory, with support from the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Funding for this work was provided by the NSF Synthetic Biology Engineering Research Center. JK received additional support via an NSF Graduate Research Fellowship.

References

1. Knight T: **Idempotent Vector Design for Standard Assembly of Biobricks.** 2003 [http://web.mit.edu/synbio/release/docs/biobricks.pdf].
2. Voigt CA: **Genetic parts to program bacteria.** *Current Opinion in Biotechnology* 2006, **17**:548-557.
3. Arkin AP, Endy D: **A Standard Parts List for Biological Circuitry.** *DARPA White Paper* 1999 [http://hdl.handle.net/1721.1/29794].
4. Endy D: **Foundations for engineering biology.** *Nature* 2005, **438**:449-453.
5. Endy D: **2003 Synthetic Biology study.** 2003 [http://hdl.handle.net/1721.1/38455].
6. Knight T: **DARPA BioComp Plasmid Distribution 1.00 of Standard Biobrick Components.** 2002 [http://hdl.handle.net/1721.1/21167].
7. Texas Instruments: *TTL logic data book: standard TTL, Schottky, low-power Schottky* Texas Instruments, Dallas, Texas, USA; 1988.
8. **Registry of Standard Biological Parts** [http://partsregistry.org]
9. Canton B, Labno A, Endy D: **Refinement and standardization of synthetic biological parts and devices.** *Nature biotechnology* 2008, **26**:787-793.
10. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Accurate prediction of gene feedback circuit behavior from component properties.** *Mol Syst Biol* 2007, **3**:143.
11. Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EPS, Landgraf D, Phillips I, Silver PA: **Rational design of memory in eukaryotic cells.** *Genes Dev* 2007, **21**:2271-2276.
12. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ: **A bottom-up approach to gene regulation.** *Nature* 2006, **439**:856-860.
13. Serebriiskii IG, Golemis EA: **Uses of lacZ to Study Gene Function: Evaluation of [beta]-Galactosidase Assays Employed in the Yeast Two-Hybrid System.** *Analytical Biochemistry* 2000, **285**:1-15.
14. Miller JH: *Experiments in Molecular Genetics* Cold Spring Harbor Laboratory, USA; 1972.
15. Eustice DC, Feldman PA, Colberg-Poley AM, Buckery RM, Neubauer RH: **A sensitive method for the detection of beta-galactosidase in transfected mammalian cells.** *BioTechniques* 1991, **11**:739-740. 742-733-739-740, 742-733
16. Pollack A: **Custom-Made Microbes at Your Service.** *New York Times. New York City* 2006.
17. **Coefficient of thermal expansion** [http://en.wikipedia.org/wiki/Coefficient_of_thermal_expansion]
18. Barnett E, Nordin BE: **The radiological diagnosis of osteoporosis: a new approach.** *Clinical Radiology* 1960, **11**:166-174.
19. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science (New York, NY)* 1995, **270**:467-470.
20. Alper H, Fischer C, Nevoigt E, Stephanopoulos G: **Tuning genetic control through promoter engineering.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:12678-83.
21. Weiss R: **Cellular computation and communications using engineered genetic regulatory networks.** 2001 [http://hdl.handle.net/1721.1/8228]. Massachusetts Institute of Technology
22. Endy : **Adventures in synthetic biology.** *Nature* 2005, **438**:449-453.
23. Leveau JHJ, Lindow SE: **Predictive and Interpretive Simulation of Green Fluorescent Protein Expression in Reporter Bacteria.** *Journal of Bacteriology* 2001, **183**:6752-62.

24. Adams CW, Hatfield GW: **Effects of promoter strengths and growth conditions on copy number of transcription-fusion vectors.** *J Biol Chem* 1984, **259:**7399-7403.
25. Shetty R, Endy D, Knight T: **Engineering BioBrick vectors from BioBrick parts.** *Journal of Biological Engineering* 2008, **2:**5-5.
26. Hawley DK, McClure WR: **Compilation and analysis of Escherichia coli promoter DNA sequences.** *Nucleic acids research* 1983, **11:**2237-2255.
27. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN: **Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays.** *Proceedings of the National Academy of Sciences* 2002, **99:**9697-9702.
28. **OpenWetWare: FACS protocol for multi-institution promoter activity experiment** [http://openwetware.org/index.php?title=The_BioBricks_Foundation:Standard_s/Technical/Measurement/Promoter_characterization_experiment_%28FACS%29&oldid=198396]
29. Lutz R, Bujard H: **Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.** *Nucleic acids research* 1997, **25:**1203-1210.
30. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R: **A synthetic multicellular system for programmed pattern formation.** *Nature* 2005, **434:**1130-1134.
31. **International Genetically Engineered Machines (iGEM) Competition** [http://igem.org]
32. Bremer H, Dennis P: **Modulation of chemical composition and other parameters of the cell by growth rate.** In *Escherichia coli and Salmonella Volume* 2. 2nd edition. Edited by: Neidhart FC. Washington DC: ASM Press; 1996:1553-1569.
33. Huxley J: *Problems of Relative Growth* Methuen, London; 1932.
34. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U: **Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria.** *Science* 2001, **292:**2080-2083.
35. Setty Y, Mayo AE, Surette MG, Alon U: **Detailed map of a cis-regulatory input function.** *Proceedings of the National Academy of Sciences* 2003, **100:**7702-7707.
36. Ronen M, Rosenberg R, Shraiman BI, Alon U: **Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics.** *Proceedings of the National Academy of Sciences* 2002, **99:**10555-10560.

Appendix 5: Publication in the Journal IET Synthetic Biology

J. Aleksic, F. Bizzari, Y. Cai, B. Davidson, K. de Mora, S. Ivakhno, S.L. Seshasayee, J. Nicholson, J. Wilson, A. Elfick, C. French, L. Kozma-Bognar, H. Ma and A. Millar

Development of a novel biosensor for the detection of arsenic in drinking water

J. Aleksic, F. Bizzari, Y. Cai, B. Davidson, K. de Mora, S. Ivakhno, S.L. Seshasayee, J. Nicholson, J. Wilson, A. Elfick, C. French, L. Kozma-Bognar, H. Ma and A. Millar

Abstract: We sought to develop a whole-cell biosensor for the detection of arsenic in drinking water, a major problem in Bangladesh and West Bengal. In contrast to previously described systems, our biosensor would give a pH change as output, allowing simple detection with a pH electrode or pH indicator solution. We designed and modelled a system based on the arsenate-responsive promoter of the *Escherichia coli* arsenic detoxification system, using urease to increase pH in the absence of arsenate, and β-galactosidase (LacZ) to decrease pH in the presence of arsenate. The pH-reducing β-galactosidase part of the system was constructed and tested, and was found to give a clear response to arsenate concentrations as low as 5 ppb arsenic, well below the World Health Organisation (WHO) recommended limit of 10 ppb.

1 Initial project ideas

In our initial discussions, we considered a variety of projects, including a biosensor for detection of water contamination, and a hybrid biological-electrical device such as a variable resistor. Ultimately we decided to combine these two ideas to develop a whole cell biosensor that responds to arsenic by producing a measurable pH change which can be easily detected with a pH electrode.

2 Usefulness of a biosensor for arsenic in drinking water

Arsenic contamination in drinking water is a serious problem in many parts of the world, and is particularly associated with Bangladesh and West Bengal, where many tube wells were inadvertently drilled through arsenic bearing sediments, resulting in drinking water contaminated with arsenate (AsO_4^3) and arsenite (AsO_3^3) anions [1–3]. Consumption of water with elevated arsenic levels over a prolonged period leads to arsenicosis, resulting in skin lesions and various cancers. Many millions of people worldwide are at risk. The current WHO recommended limit for drinking water is 10 ppb arsenic; in many countries a more relaxed limit of 50 ppb is still in operation. Current field tests for arsenic detection are not altogether satisfactory, requiring toxic chemicals and having a significant false negative rate.

A simple, cheap and sensitive field assay for arsenic levels would therefore be extremely useful. A whole cell microbial biosensor, with an arsenic-responsive promoter linked to a suitable reporter gene, might be one way of achieving this [4, 5]. Arsenic biosensors have been previously reported [for example, 6, 7], but have mainly relied on luminescent or fluorescent reporter genes, which require expensive equipment and trained technicians, and are not really suitable for field use. Other biosensors have used the LacZ/Xgal reporter system, but this is difficult to quantify, and Xgal (5-bromo-4-chloro-3-indoyl-ß-D-galactoside) is expensive and requires refrigeration. By contrast, a sensor giving a pH response would allow a simple quantitative measurement using a cheap pH electrode or solid state device (ISFET, ion-sensitive field effect transistor), or even just a pH indicator solution giving a colour change.

3 Design of the system

We devised a system based on the plasmid-encoded arsenic resistance operon of *Escherichia coli*. This is controlled by two repressor proteins, ArsR (responding to low concentrations of arsenate or arsenite) and ArsD (responding to higher concentrations)[8, 9]. Each is negatively autoregulated. To induce an increase in pH, we chose to use urease, which breaks down urea, $(NH_2)_2CO$, to release ammonium ions. This is used in diagnostic microbiology to distinguish urease-positive bacteria such as *Proteus*, since the pH can rise above 9. To induce a decrease in pH, we chose to use *lacZ*. This encodes β-galactosidase, which catalyses the essential first step in the fermentation of lactose to acetic and lactic acids (mixed acid fermentation) in *E. coli* and related organisms. This reaction is also used in diagnostic microbiology, since the pH can fall below 4.5.

In our design, the activity of the biosensor is initiated by exposure to lactose. Urease is expressed from a hybrid promoter repressed by both lambda cI repressor and LacI repressor. In the presence of lactose, but absence of arsenate, urease is induced and the pH rises. When low amounts of arsenate are present, an ArsR-repressed promoter is induced, leading to expression of lambda cI repressor, switching off urease production. Thus the pH remains neutral. If higher amounts of arsenate are present, *lacZ* expression is induced through an ArsD-responsive

doi:10.1049/iet-stb:20060002

Paper first received 30th December 2006

J. Aleksic, F. Bizzari, J. Nicholson, C. French, L. Kozma-ognar and A. Millar are with the School of Biological Sciences, University of Edinburgh, Edinburgh, UK

B. Davidson, K. de Mora, J. Wilson and A. Elfick are with the School of Engineering and Electronics, University of Edinburgh, Edinburgh, UK

Y. Cai, S. Ivakhno, S.L. Seshasayee and H. Ma are with the School of Informatics, University of Edinburgh, Edinburgh, UK

E-mail: c.french@ed.ac.uk

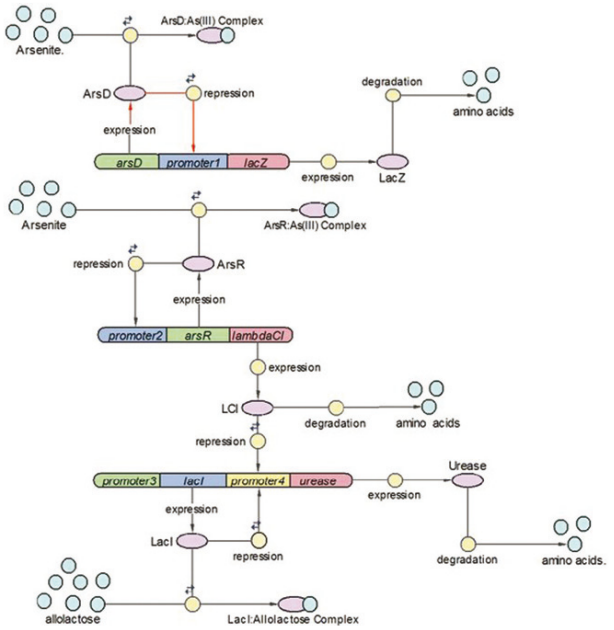


Fig. 1 *Schematic diagram of the complete biosensor system*

promoter, leading to a fall in pH. By using multiple promoters in this way, a high sensitivity and high dynamic range (the range of arsenate concentrations over which arsenate concentration can be estimated from the response) are achieved. A schematic diagram of the system is shown in Fig. 1.

4 Modelling

This system was modelled using an ordinary differential equation (ODE)-based model, with parameters estimated based on the literature. The model showed good induction of urease and repression of *lacZ* in the absence of arsenate, and repression of urease and induction of *lacZ* at high arsenate levels. Sensitivity analysis was also conducted in order to determine which parameters had the greatest effect on the urease and *lacZ* responses. For example, it was found that the parameter having the greatest effect on the steady-state level of *lacZ* expression in the presence of arsenate was the degradation rate of ArsD.

5 Testing the concept

To demonstrate that a detectable pH change could be achieved in the laboratory, we constructed a BioBrick bearing the *E. coli* chrososomal *ars* promoter and negatively autoregulated *arsR* gene (BBa_J33201). The chromosomal *ars* operon is similar to the plasmid-encoded one we had originally envisaged using, but is controlled solely by ArsR and has no equivalent of the second repressor, ArsD [10, 11]. We also made a BioBrick (BBa_J33202) of the *lacZ′* gene encoding the N-terminus of *lacZ*, which complements the *lacZ*ΔM15 mutation found on the chromosome of laboratory strains of *E. coli* such as JM109 and XL1Blue. These BioBricks were joined to generate BBa_J33203. Unfortunately, we were not able to obtain template DNA for the plasmid encoded *arsR* and *arsD* genes we had intended to use within the time frame of the competition. We therefore also cloned the *ars* promoter and *arsR* gene from *Bacillus subtilis* [12], to test whether this might have a sufficiently different affinity for arsenate to be useful in this context. This was joined to *lacZ′* to generate BBa_J33206. In experiments using JM109/pSB1A2-BBa_J33203, concentrations of arsenate as low as 5 ppb gave a significant decrease in pH at incubation times above 5 h, persisting to over 20 h, in a non-optimised medium based on Luria-Bertani medium with 2% w/v lactose (Fig. 2). The response was easily detected with a pH electrode and could also be visually assessed using the pH indicator methyl red, which has a pKa around 4.8 5.0. The equivalent system using BBa_J33206 unfortunately did not show a response to arsenate, with even arsenate-free controls giving a rapid drop in pH, suggesting

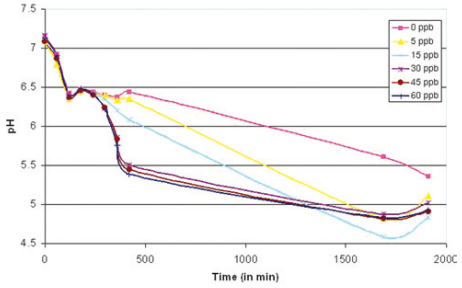


Fig. 2 *pH response of E. coli JM109/pSB1A2-BBa_J33203 to varying arsenate concentrations*

Fig. 3 *The University of Edinburgh iGEM2006 team from left to right: Jennifer Wilson; Kim de Mora;Bryony Davidson; Jelena Aleksic; Judith Nicholson; Sergii Ivakhno; Sreemati LalgudiSeshasayee; Dr Laszlo Kozma-Bognar; Yizhi Cai; Dr Chris French; Dr Hongwu Ma; Dr Alistair Elfick; Farid Bizzari; Prof. Andrew Millar*

high background activity due to incomplete repression of this promoter in *E. coli*.

6 Generation of other parts

For our system, we also needed a urease part to increase pH in the absence of arsenate. The most obvious choice would have been the urease gene cluster present in some strains of *E. coli*; however, this consists of around 5 kb of DNA with seven genes (*ureDABCEFG*, where *ureABC* are the genes encoding the urease subunits, and the other genes encode accessory factors required for proper insertion of the nickel cofactor) and contains six forbidden restriction sites which would have to be mutated out individually before the gene cluster could be converted to a BioBrick. After searching the literature, we found that the *Bacillus subtilis* urease gene cluster consists of only three genes, *ureABC*, which can nevertheless be assembled into a functioning urease in *E. coli* without the requirement for the usual accessory proteins [13]. Unfortunately, this gene cluster also contains two forbidden restriction sites, EcoRI and SpeI.

To check that this urease would be suitable for our purposes, the *ureABC* region was cloned in pGemT-easy (Promega) and pBluescript SK+ (Stratagene). In both constructs, activity was demonstrated in *E. coli*, with pH rising to 9 after incubation in LB with 0.2% w/v urea. Having decided that this urease would be suitable, we used site-directed mutagenesis to remove the two forbidden restriction sites. This was successfully achieved, but the mutant gene cluster gave no detectable urease activity. Sequencing revealed a possible frameshift mutation in *ureC* as well as two non-silent single nucleotide changes as compared to the published sequence. Thus we were unable to generate a urease BioBrick during the time available.

The final part required for our system was the hybrid promoter repressed by both lambda cI and LacI. This was generated by fusing the P_{RM}-P_R region of bacteriophage lambda, including cI binding sites OR1, OR2 and OR3, to the 3′ end of the lac promoter region including the LacI binding site. The N-terminal region of *lacZ* was also included, so that *lacZ′* expression could be used to test regulation of the promoter. This BioBrick was designated BBa_J33205. Unfortunately, we did not have time to build the constructs necessary to test the regulation of this part.

Fig. 4 *The University of Edinburgh iGEM 2006 team logo*

7 Conclusions

Even though we were not able to build our complete design in the time available, we have demonstrated that a simpler version, *E. coli* JM109/pSB1A2-BBa_J33203, gives a good pH response to arsenate concentrations as low as 5 ppb arsenic, with a dynamic range in the region of 0 20 ppb, in a non-optimised system. This can be detected with a pH electrode or a pH indicator (methyl red) which

changes from yellow to red when the pH falls below about 5. Recalling that the WHO limit is 10 ppb, this device is suitable for further development, and could potentially be the basis for a cheap and useful sensor to help prevent the ongoing tragedy of chronic arsenic poisoning. Also, we have submitted the functioning arsenic-responsive promoter to the Registry (BBa_J33201), so we hope that others may be inspired to develop even better arsenic biosensors in the future.

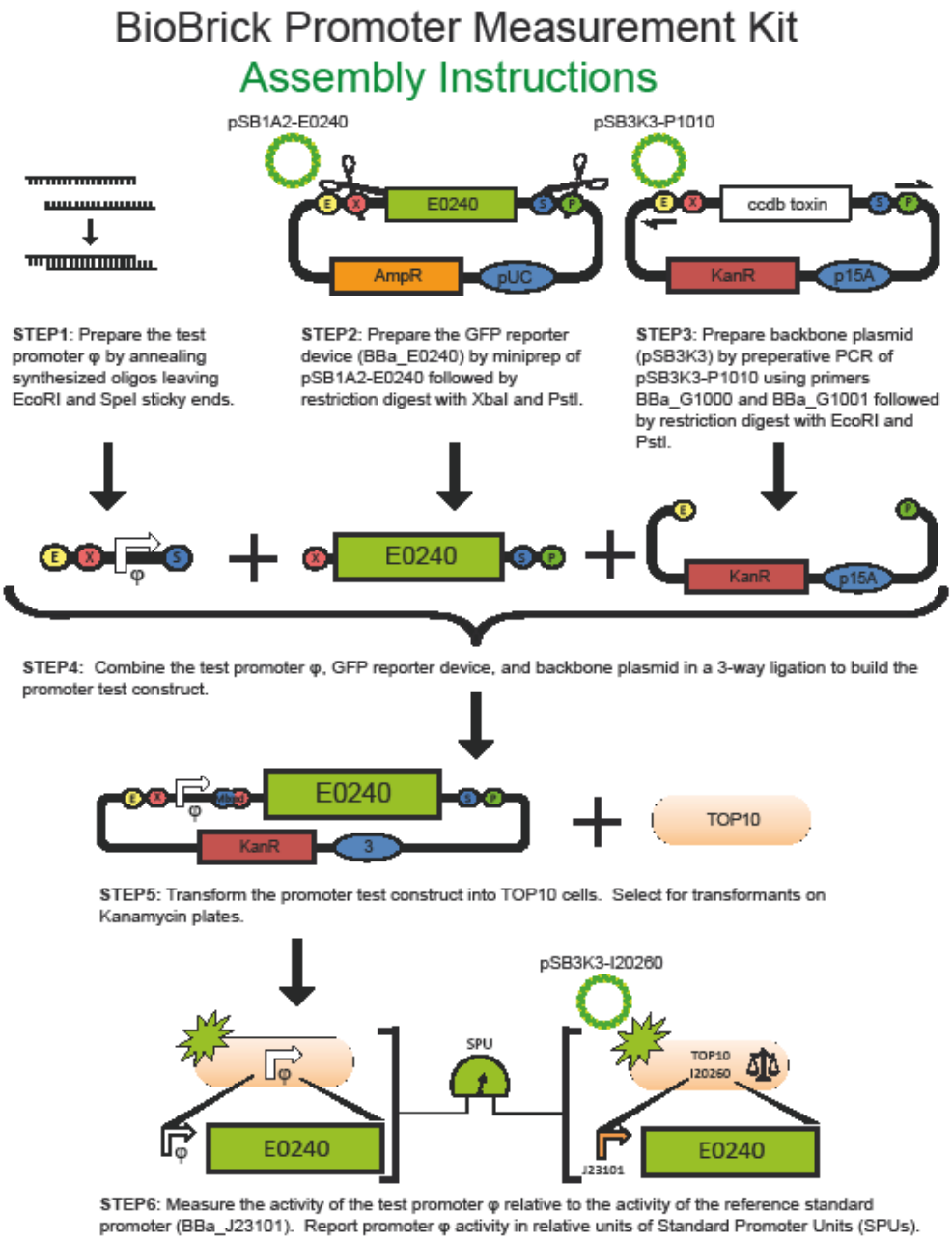
8 Acknowledgments

The Edinburgh iGEM2006 team, Team Macteria (Figs. 3 and 4), gratefully acknowledges financial support from the Gatsby Foundation, the Royal Commission for the Exhibition of 1851, SYNBIOCOMM, the Biotechnology and Biological Sciences Research Council, and the Engineering and Physical Sciences Research Council. We would also like to thank Mathworks for allowing us to use the Simbiology software for mathematical modelling of our system.

9 References

1 Meharg, A.: 'Venomous Earth: how arsenic caused the world's worst mass poisoning', Macmillan Publishers, 2005
2 Chowdhury, A.M.R.: 'Arsenic crisis in Bangladesh', *Sci. Amer.*, 2004, **291**, (2), pp. 71–75
3 Tareq, S.M., Safiullah, S., Anawar, H.M., Rahman, M.M., and Ishizuka, T.: 'Arsenic pollution in groundwater: a self-organizing complex geochemical process in the deltaic sedimentary environment', *Sci. Total Environ.*, 2003, **313**, pp. 213–226
4 Belkin, S.: 'Microbial whole-cell sensing systems of environmental pollutants', *Curr. Opin. Microbiol.*, 2003, **6**, pp. 206–212
5 Daunert, S., Barret, G., Feliciano, J.S., Shetty, R.S., Shrestha, S., and Smith-Spencer, W.: 'Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes', *Chem. Rev.*, 2000, **100**, pp. 2705–2738
6 Tauriainen, S., Karp, H., Chang, W., and Virta, M.: 'Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite', *Appl. Environ. Microbiol.*, 1997, **63**, pp. 4456–4461
7 Stocker, J., Balluch, D., Gsell, M., Harms, H., Feliciano, J., Daunert, S., Malik, K.A., and Van der Meer, J.R.: 'Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water', *Environ. Sci. Technol.*, 2003, **37**, pp. 4734–4750
8 Wu, J.H., and Rosen, B.P.: 'The *arsD* gene encodes a second trans-acting regulatory protein of the plasmid-encoded arsenical resistance operon', *Mol. Microbiol.*, 1993, **8**, pp. 615–623
9 Chen, Y.-X., and Rosen, B.P.: 'Metalloregulatory properties of the ArsD repressor', *J. Biol. Chem.*, 1997, **272**, pp. 14257–14262
10 Diorio, C., Cai, J., Marmor, J., Shinder, R., and DuBow, M.S.: 'An *Escherichia coli* chromosomal *ars* operon homolog is functional in arsenic detoxification and is conserved in Gram negative bacteria', *J. Bacteriol.*, 1995, **177**, pp. 2050–2056
11 Cai, J., and DuBow, M.S.: 'Expression of the *Escherichia coli* chromosomal *ars* operon', *Canad. J. Microbiol.*, 1996, **42**, pp. 662–671
12 Sato, T., and Kobayashi, Y.: 'The *ars* operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite', *J. Bacteriol.*, 1998, **180**, pp. 1655–1661
13 Kim, J.-K., Mulrooney, S.B., and Hausinger, R.P.: 'Biosynthesis of active *Bacillus subtilis* urease in the absence of known urease accessory proteins', *J. Bacteriol*, 2005, **187**, pp. 7150–7154

Appendix 6a: BioBrick Promoter Measurement Kit Assembly Instructions



Appendix 6b: BioBrick Promoter Measurement Kit Measurement Instructions

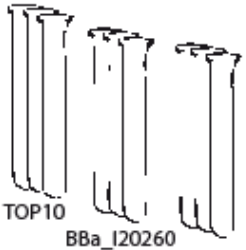
BioBrick Promoter Measurement Kit

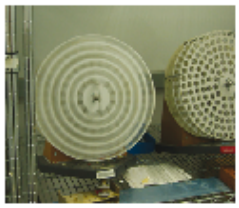
Measurement Instructions

STEP1: Streak 3 plates

A: TOP10
B: BBa_I20260
C: Your promoter!

STEP 2: Pick 3 colonies from each plate to start overnight cultures in Supplemented M9 Media at 37 C (9 tubes)

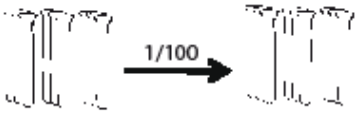




37C

16 hours

STEP 3: Dilute 1/100 into fresh, pre-warmed media incubate at 37C (9 tubes)



37C

STEP 4: After 3 hours measure GFP and OD

3 hours

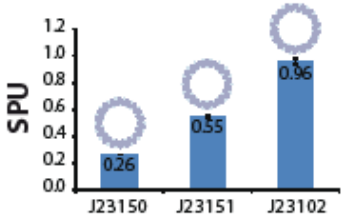
STEP 5: After another half hour measure GFP and OD again

1/2 hour

STEP 6: Input your data into the SPU Calculator and report your results at:
http://partsregistry.org/measurement

Practice Promoter Set

Weak (J23150), medium (J23151), and strong (J23102) promoters have already been assembled into promoter test constructs and are ready to measure! Transform the 3 test constructs into TOP10 and measure using the protocol above. Report your results at:
http://partsregistry.org/measurement



Appendix 7: Image Pro 7.0 Macro V.5 script

```
Public BlueId As Integer, YellowId As Integer, RedId As Integer
Public BlankId As Integer, YeastId As Integer, ProcessedYeast As
Integer
Public DoughnutId As Integer, ProcessedYellow As Integer,
InvertMask As Integer, ProcessedBlue As Integer, ProcessedRed As
Integer

' Default Script
Option Explicit
```

```
Sub StartHere
    ret = MsgBox("This is Version 5 of the macro (uses RFP image
to make mask)")
     OpenFile
     ProcessBlank_Yeast
     MeasureYFP
     ret = MsgBox("Make sure Excel is open and a new workbook
selected")
     Step1CreateOutlines
     MeasureCFP
     ret = MsgBox("Move data from YFP to sheet 2 in Excel as
sheets one will be overwritten")
     Step1CreateOutlines
     MeasureRFP
     ret = MsgBox("Move data from CFP to sheet 2 in Excel as
sheets one will be overwritten")
     Step1CreateOutlines
     ret = IpAoiManager(AOIDELETE,"")'delete all AOIs
     ret = MsgBox("Macro Finished")
End Sub
```

```
Sub OpenFile()
Dim x As Integer
     ret = MsgBox("Open Files in the order CFP, YFP, RFP")
     'Open and max intensity composite 3 images
     ret = IpTemplateMode(1)
     For x = 0 To 2
          'ret                                                =
IpWsLoad("\\129.215.237.150\s0129661\Microscopy\10_S7_1_R3D     -
470 nm.seq","seq")
          'ret        =        IpWsLoad("E:\Kim_test\10-25-09_1-
3h\10_S7_1_R3D - 470 nm.seq","seq")
```

```
            'ret                                                    =
IpWsLoad("G:\Kim_IMPro_processing\7_24h\10_S7_1_R3D     -     470
nm.seq","seq")
            ret = IpWsLoad("E:\Kim_test\12-01-09\ - 470 nm - 470
nm.seq","seq")

      Next x

BlueId = 3
YellowId =4
RedId = 5

'BlueId = 0
'YellowId =1
'RedId = 2

ret = IpBlbShow(1)
      For x = 0 To 2
            ret = IpEDFAdd(x)
            ret = IpEDFNew(x)
            ret = IpEDFSet(EDF_NORMALIZE, 0, 0)
            ret = IpEDFSet(EDF_CRITERIA, EDF_MAX_INTENSITY, 0)
            ret = IpEDFSet(EDF_ORDER, EDF_BOTTOMUP, 0)
            ret = IpEDFCreate(EDF_COMPOSITE)
      Next x
      ret = IpTemplateMode(0)
      For x = 0 To 2
            ret = IpAppSelectDoc(x)
            ret = IpDocClose()
      Next x

      'Open and prepare Blank Image
      ret = MsgBox("Open Blank Image")
      ret = IpTemplateMode(1)
      'ret                                                        =
IpWsLoad("\\129.215.237.150\s0129661\Microscopy\10_S7_1_R3D     -
470 nm.seq","seq")
    'ret = IpWsLoad("E:\Kim_test\10-25-09_1-3h\10_S7_1_R3D - 470
nm.seq","seq")
      'ret  =  IpWsLoad("G:\Kim_IMPro_processing\7_24h\10_S7_1_R3D
- 470 nm.seq","seq")
      ret = IpWsLoad("E:\Kim_test\12-01-09\10_S5-CYR_1_R3D\ - 470
nm - 470 nm.seq","seq")
      ret = IpTemplateMode(0)
      ret = IpDocGet(GETACTDOC, 0, BlankId)
      ret = MsgBox("Open yeast image")
      ret = IpTemplateMode(1)
```

```
      'ret                                                              =
IpWsLoad("\\129.215.237.150\s0129661\Microscopy\10_S7_1_R3D     -
470 nm.seq","seq")
      'ret  =  IpWsLoad("E:\Kim_test\10-25-09_1-3h\10_S7_1_R3D  -
470 nm.seq","seq")
      'ret  =  IpWsLoad("G:\Kim_IMPro_processing\7_24h\10_S7_1_R3D
- 470 nm.seq","seq")
     ret  =  IpWsLoad("E:\Kim_test\12-01-09\10_S5-CYR_1_R3D\  -  470
nm - 470 nm.seq","seq")
      ret = IpTemplateMode(0)
      ret = IpDocGet(GETACTDOC, 0, YeastId)
End Sub
```

```
Sub ProcessBlank_Yeast()

      ret = IpAppSelectDoc(BlankId)
      'ret = IpOpNumberArithmetics(3000.0, OPA_SUB, 0)
      ret  =  IpOpNumberArithmetics(1500.0,  OPA_SUB,  0)  'This  is
the line that changes the numerical subtract from the background
      ret = IpAppSelectDoc(YeastId)
      ret = IpOpShow(1)
      ret = IpOpImageArithmetics(BlankId, 0.0, OPA_SUB, 1)
      ret = IpOpShow(0)
      ret = IpDocGet(GETACTDOC, 0, ProcessedYeast)
      ret = IpAppSelectDoc(YeastId)
      ret = IpDocClose()
      ret = IpAppSelectDoc(BlankId)
      ret = IpDocClose()

      ret = IpAppSelectDoc(BlankId)
      ret = MsgBox("Select part of stack to best show doughnuts")
      ret = IpEDFAdd(ProcessedYeast)
      ret = IpEDFNew(ProcessedYeast)
      ret = IpEDFSet(EDF_NORMALIZE, 0, 0)
      ret = IpEDFSet(EDF_CRITERIA, EDF_MAX_INTENSITY, 0)
      ret = IpEDFSet(EDF_ORDER, EDF_BOTTOMUP, 0)
      ret = IpEDFCreate(EDF_COMPOSITE)
      ret = IpDocGet(GETACTDOC, 0, DoughnutId)

      ret = IpAppSelectDoc(ProcessedYeast)
      ret = IpDocClose()

      ret = IpAppSelectDoc(DoughnutId)
      ret = IpFltLoPass(3, 10, 5)
      ret = IpSegSetRange(0, 2048, 65535)
      ret  =  MsgBox("Threshold  to  include  as  many  doughnuts  as
possible")
```

```
      ret     =     IpBlbSetFilterRange(BLBM_AREA,     307.0921021,
30709.21094)
      ret = IpBlbCount()
      ret = IpBlbUpdate(0)
      ret = IpBlbCreateMask()


      '*************************************************************
***
      '**                                  FILTERING SECTION
                  **
      '*************************************************************
***
      ret = IpLSFltApply(LF_HIPASS,10,10,2,1)     'Large  Spectral
HighPass
      'ret = MsgBox("Manually Threshold images to select halos")
      'ret = IpBlbCount()
      'ret = IpBlbUpdate(0)
      'ret = IpBlbCreateMask()
      ret = IpFltClose(MORPHO_2x2SQUARE, 3) 'Close Filter
      ret = IpFltDilate(MORPHO_2x2SQUARE, 3)'Dilate Filter
      ret = IpFltErode(MORPHO_2x2SQUARE, 3) 'Erode Filter
      ret = IpFltPrune(50, -1)                    'Prune Filter

      '*************************************************************
***
      '**                     END OF FILTERING SECTION
               **
      '*************************************************************
***

      'Produce Yeast cell AOI
      ret     =     IpBlbSetFilterRange(BLBM_AREA,     153.5460510,
2456.736816)  ' Area filter range of 2.5 - 40
      'ret = IpBlbSetAttr(BLOB_BRIGHTOBJ, 0)      'Turn         on
measure dark objects
      ret = MsgBox("Select Automatic Dark Objects")
      ret = IpBlbCount()                                          '
Count dark objects
      ret = IpBlbUpdate(0)
      ret                                                         =
IpBlbSaveOutline("\\129.215.237.150\s0129661\Microscopy\Outlines
_Temp\Outlines.scl")  'Save resulting cell outlines

      ret = IpDocClose()
      ret = IpAppSelectDoc(DoughnutId)
      ret = IpDocClose()
End Sub
```

```
 Sub MeasureYFP
      Dim RedIdDup As Integer, RedIdSub As Integer

      ret = IpAppSelectDoc(RedId) 'Select RFP image
      ret = IpWsDuplicate()       'Duplicate RFP image (one for
low pass filtering
      ret = IpDocGet(GETACTDOC, 0, RedIdDup)
      ret = IpLSFltApply(LF_LOPASS,10,10,2,1)     'Large Spectral
HighPass
      ret = IpAppSelectDoc(RedId)
      ret = IpOpImageArithmetics(RedIdDup, 0.0, OPA_SUB, 1)
      'Subtract LoPassed duplicate image from original RFP image
      ret = IpDocGet(GETACTDOC, 0, RedIdSub)
      ret = IpAppSelectDoc(RedIdSub)              'Select
subtracted RFP image
      ret = IpLSFltApply(LF_HIPASS,10,10,2,1)     'Apply highpass
filter
      'ret    =    IpBlbSetFilterRange(BLBM_AREA,    6.141842365,
12.28368473)    'Set area measurement filters 0.1 -0.2
      'ret    =    IpBlbSetFilterRange(BLBM_AREA,    9.212763786,
73.70211029)    'Set area measurement filters 0.15 - 1.2
      'ret = IpBlbSetFilterRange(BLBM_AREA,  5.86, 46.92)
    ret = IpBlbSetFilterRange(BLBM_AREA, 10, 70) 'Set       area
measurement filters ~0.15 - 1.2
      ret = IpSegShow(1)
      ret = IpSegSetRange(0, 2048, 65535)
      ret = MsgBox("Set Threshold Values")
      ret = IpBlbCount()              'Count Objects in image
      ret = IpBlbUpdate(0)
      ret = IpBlbCreateMask()         'Make   mask   of   counted
objects
      ret = IpLutSetAttr(LUT_CONTRAST, -2)    'Invert Mask
      ret = IpDocGet(GETACTDOC, 0, InvertMask)
      ret = IpAppSelectDoc(YellowId)   'Select the YFP image
      ret = IpOpImageArithmetics(InvertMask, 0.0, OPA_SUB, 1)
      ret = IpDocGet(GETACTDOC, 0, ProcessedYellow)
      ret = IpSegSetRange(0, 257, 65535)
      ret = IpSegPreview(CURRENT_C_T)
      ipICal(0) = 257
      ipICal(1) = 65535
      ret = IpBlbMultiRanges(ipICal(0), 1)

      'ret = IpAppSelectDoc(RedId)
      'ret = IpDocClose()
      ret = IpAppSelectDoc(RedIdSub)
      ret = IpDocClose()
```

```
End Sub
```

```
Sub MeasureCFP
      ret = IpAppSelectDoc(BlueId)     'Select the CFP image
      ret = IpOpImageArithmetics(InvertMask, 0.0, OPA_SUB, 1)
      ret = IpDocGet(GETACTDOC, 0, ProcessedBlue)
      ret = IpSegSetRange(0, 257, 65535)
      ret = IpSegPreview(CURRENT_C_T)
      ipICal(0) = 257
      ipICal(1) = 65535
      ret = IpBlbMultiRanges(ipICal(0), 1)
End Sub
```

```
Sub MeasureRFP
      ret = IpAppSelectDoc(RedId)'Select the RFP image
      ret = IpOpImageArithmetics(InvertMask, 0.0, OPA_SUB, 1)
      ret = IpDocGet(GETACTDOC, 0, ProcessedRed)
      ret = IpSegSetRange(0, 257, 65535)
      ret = IpSegPreview(CURRENT_C_T)
      ipICal(0) = 257
      ipICal(1) = 65535
      ret = IpBlbMultiRanges(ipICal(0), 1)
End Sub
```

```
'Count objects on image 1 and then execute the macro
Sub Step1CreateOutlines
      ret  =  IpMacroStop("Count  objects  on  image  1  .  Click
Continue to save the outlines of every object to AOI manager.
Then go to step 2.", 0)
      If ret=2 Then Exit Sub    'cancel
      ReDim blbpts(1000) As POINTAPI
      Dim numpoints As Integer, numobj As Integer
      Dim status As Integer, i As Integer
      ' get the total number of objects, in-range and out-of-
range,
      ' hidden and visible.

'ret = IpAppSelectDoc(ProcessedYellow)
ret = IpBlbSetFilterRange(BLBM_AREA, 153.5460510, 2456.736816)
ret                                                          =
IpBlbLoadOutline("\\129.215.237.150\s0129661\Microscopy\Outlines
_Temp\Outlines.scl")

      ret = IpBlbGet(GETNUMOBJEX, 0, 0, numobj)
      IpAoiManager(AOIDELETE,"")'delete all AOIs
      For i = 0 To numobj - 1
         numpoints = IpBlbGet(GETPOINTS, i, 1000, blbpts(0))
```

```
         If numpoints > 0 Then
              ' create AOI out of the object outline and XOR it.
              ret = IpAoiCreateIrregular(blbpts(0), numpoints)
                 ret = IpAoiManager(AOIADD, "Polygon" & Str(i+1))
         End If
      Next i
      IpIniFile(INICMD_SETINT,"NumberOfAOIs",numobj)
      ret = IpAoiShow(FRAME_NONE)
      Count
End Sub
```

```
Sub Count()
      ret = IpMacroStop("Set threshold and count/size parameters
on image 2 and click Continue to measure object parameters
withing saved on step 1outlines .", 0)
      If ret=2 Then Exit Sub'cancel
      Dim i As Integer, numobj As Integer, status As Integer,
numeasure As Integer, x As Integer
Dim FirstCol As String, SecondCol As String, ThirdCol As String,
FourthCol As String, FifthCol As String
Dim Row As Integer

      ret=IpIniFile(INICMD_GETINT,"NumberOfAOIs",numobj)
      ' Open communication with sheet1 of Excel
      ret = IpDde(DDE_OPEN, "Excel", "sheet1")
      ret = IpDde(DDE_PUT, "R1C1", "Area")
      ret = IpDde(DDE_PUT, "R1C2", "Roundness")
      ret = IpDde(DDE_PUT, "R1C3", "Max Intensity")
      ret = IpDde(DDE_PUT, "R1C4", "Average Intensity")
      ret = IpDde(DDE_PUT, "R1C5", "Cell Number")

      'set append data to the bottom
      ret = IpDde(DDE_SET, "row", "3")
      'ret = IpDde(DDE_SET, "append", "1")
      ret = IpDde(DDE_SET, "row", "1")
      Row = 2
ret = IpBlbEnableMeas(BLBM_DENSITY, 1)
ret = IpBlbEnableMeas(BLBM_DENSMAX, 1)
ret = IpBlbEnableMeas(BLBM_ROUNDNESS, 1)
'ret = IpBlbSetFilterRange(BLBM_AREA, 6.141842365, 12.28368473)
ret = IpBlbSetFilterRange(BLBM_AREA, 10, 70)     'Set        area
measurement filters ~0.15 - 1.2
For i=0 To numobj-1
      ret = IpAoiManager(AOISET, "Polygon" & Str(i+1))
      ret = IpBlbCount()'count within AOI
      ret = IpBlbUpdate(0)
      'export data to Excel appending the data to the bottom
      If i=0 Then
```

```
            '
      ret=IpBlbSaveData("",S_Y_AXIS+S_HEADER+S_DDE)'first    block
with header
      Else


                  ret = IpBlbGet(GETNUMOBJEX, 0, 0, numeasure)
                  Dim PutVal As Single
                  For x = 0 To numeasure -1
                        ret = IpBlbGet(GETSTATUS, x, 0, status)

                        ' Put values into cell on appropriate row
and column.

                  If status >= 0 Then
                        FirstCol = "R" + CStr(Row) + "C1"
                        ret = IpBlbData(BLBM_AREA, x, x, PutVal)
                        ret      =      IpDde(DDE_PUT,      FirstCol,
Str(PutVal))

                        SecondCol = "R" + CStr(Row) + "C2"
                        ret   =   IpBlbData(BLBM_ROUNDNESS,   x,   x,
PutVal)
                        ret      =      IpDde(DDE_PUT,      SecondCol,
Str(PutVal))

                        ThirdCol = "R" + CStr(Row) + "C3"
                        ret   =   IpBlbData(BLBM_DENSMAX,   x,   x,
PutVal)
                        ret      =      IpDde(DDE_PUT,      ThirdCol,
Str(PutVal))

                        FourthCol = "R" + CStr(Row) + "C4"
                        ret   =   IpBlbData(BLBM_DENSITY,   x,   x,
PutVal)
                        ret      =      IpDde(DDE_PUT,      FourthCol,
Str(PutVal))

                        FifthCol = "R" + CStr(Row) + "C5"
                        ret = IpDde(DDE_PUT, FifthCol, Str(i))
                        Row = Row + 1
                  End If
            Next x
            'ret  =  IpBlbSaveData("",    S_Y_AXIS+S_DDE)'the  rest
without headers
      End If
Next i
```

```
End Sub
```

Appendix 8: Matlab image processing script

```
% This script is for processing Macro V5 spreadsheets
% Scripts needed to run MacroV5:
%
%   Labelmaker
%   Structuremaker
%   XLParse
%   Structuremaker2
%   CYRStructuremaker
%   Plot_figure1
%   Plot_figure2
%   KJY15_Fit
%   Plottest2

clc
clear
format bank

% File to load

%FileName = '2-12-09_KJY15_Macro_V5.xls';
%FileName = 'KJY10_MacroV5_12-01-09';  %Works!
FileName = 'KJY11_MacroV5_12-01-09';  %Works!
%FileName = 'KJY14_MacroV5_12-01-09_2'; %Works minus one image
%FileName = 'KJY15_MacroV5_12-01-09'; %Works!

Data = [];
%This script creates all the necessary labels
%-------------------------------------------------------------------------
----

Labels = [];
Labels.YFPlabel = 1;
Labels.CFPlabel = 2;
Labels.RFPlabel = 3;

[XLnum XLtxt] = xlsread(FileName);

Labels.ChannelNameCol = 1;
Data.ChannelLabels = XLnum(:, Labels.ChannelNameCol);

Labels.AreaCol = 2;
Data.Areas = XLnum(:, Labels.AreaCol);

Labels.RoundnessCol = 3;
Data.Roundness = XLnum(:, Labels.RoundnessCol);
```

```
Labels.MaxIntensityCol = 4;
Data.MI = XLnum(:, Labels.MaxIntensityCol);

Labels.AverageIntensityCol = 5;
Data.AI = XLnum(:, Labels.AverageIntensityCol);

Labels.CellNumberCol = 6;
Data.CellNumber = XLnum(:, Labels.CellNumberCol);
%-------------------------------------------------------------------------
----
% %Parse out YFP, CFP and RFP Channels;

%The Arraymaker script conatins all the variables for the CFP, YFP and
RFP
%structures
%-------------------------------------------------------------------------
----
YFP.AI = [];
CFP.AI = [];
RFP.AI = [];
YFP.Areas = [];
CFP.Areas = [];
RFP.Areas = [];
YFP.Roundness = [];
CFP.Roundness = [];
RFP.Roundness = [];
YFP.MI = [];
CFP.MI = [];
RFP.MI = [];
YFP.CellNumber = [];
CFP.CellNumber = [];
RFP.CellNumber = [];
YFP.ImageAverage = [];
CFP.ImageAverage = [];
RFP.ImageAverage = [];
YFP.ImageMultiplier = [];
CFP.ImageMultiplier = [];
YFP.ImageAdjust = [];
CFP.ImageAdjust = [];
YFP.ImageNorm1 = [];
CFP.ImageNorm1 = [];
CFP.AIMean = [];
YFP.AIMean = [];
RFP.AIMean = [];
CFP.AIMeanArray = [];
YFP.AIMeanArray = [];
RFP.AIMeanArray = [];
CFP.AINormmean = [];
YFP.AINormmean = [];
RFP.AIMean2 = [];
CFP.AIMean2 = [];
```

```
YFP.AIMean2 = [];
RFP.AINorm2 = [];
CFP.AINorm2 = [];
YFP.AINorm2 = [];
CFP.CFP2 = [];
YFP.YFP2 = [];
RFP.RFP2 = [];
CFP.CFP3 = [];
YFP.YFP3 = [];
RFP.RFP3 = [];
CFP.CFP4 = [];
YFP.YFP4 = [];
RFP.RFP4 = [];

Results = [];
%-------------------------------------------------------------------------
----

% This section writes the Excel YFP, CFP and RFP data to the structures
%-------------------------------------------------------------------------
----
% This .m file is part of the MacroV5 parse script.  It writes the
XLnum data
%into the correct locations in the (X)FP.structure.  It was made into a
% separate file to reduce the length of MacroV5.

for i = 1:length(Data.AI);

   if isnan(Data.AI(i))
       i;

       continue
   end

   if Labels.YFPlabel == Data.ChannelLabels(i);
       YFP.AI = [YFP.AI Data.AI(i)];
       YFP.Areas = [YFP.Areas Data.Areas(i)];
       YFP.Roundness = [YFP.Roundness Data.Roundness(i)];
       YFP.MI = [YFP.MI Data.MI(i)];
       YFP.CellNumber = [YFP.CellNumber Data.CellNumber(i)];
   end

   if Labels.CFPlabel == Data.ChannelLabels(i);
       CFP.AI = [CFP.AI Data.AI(i)];
       CFP.Areas = [CFP.Areas Data.Areas(i)];
       CFP.Roundness = [CFP.Roundness Data.Roundness(i)];
       CFP.MI = [CFP.MI Data.MI(i)];
       CFP.CellNumber = [CFP.CellNumber Data.CellNumber(i)];
   end

   if Labels.RFPlabel == Data.ChannelLabels(i);
```

```
        RFP.AI = [RFP.AI Data.AI(i)];
        RFP.Areas = [RFP.Areas Data.Areas(i)];
        RFP.Roundness = [RFP.Roundness Data.Roundness(i)];
        RFP.MI = [RFP.MI Data.MI(i)];
        RFP.CellNumber = [RFP.CellNumber Data.CellNumber(i)];
    end

     if Labels.YFPlabel ~= Data.ChannelLabels(i) | Labels.CFPlabel ~=
Data.ChannelLabels(i) | Labels.RFPlabel ~= Data.ChannelLabels(i);
         %i; "~=" is the NOT EQUALS operator
     end

end
%-------------------------------------------------------------------------
----

%-%Exclude Zeros%
BadData = 1;
BadData = [];
BadData = union(BadData, find(log(YFP.AI ./ CFP.AI) < -3));
BadData = union(BadData, find(YFP.AI == 0));

CFP.AI(BadData) = [];
YFP.AI(BadData) = [];
CFP.Areas(BadData) = [];
YFP.Areas(BadData) = [];

%This section is about normalizing the CFP and the YFP data to the
image to
%image average and uses the RangeN array

StartPos = 1;
ImageNumber = 1;
RangeN = [];

for i = 1:length(Data.AI);

    if isnan(Data.AI(i))
        i;

        FinPos=i-1;
        RangeN = [RangeN; StartPos,...
                  FinPos,...
                  (FinPos - StartPos+1),...
                  Data.CellNumber(FinPos),...
                  mean(Data.AI(StartPos:FinPos))];
        XLnum(StartPos:FinPos, 7) = RangeN(ImageNumber, 5);
        StartPos=i+1;
        ImageNumber=ImageNumber+1;
        continue
    end
```

```
end

%Writing the ImageAverage data to the XLnum array
Labels.ImageAverageCol = 7;
Data.ImageAverage = XLnum(:, Labels.ImageAverageCol);

% This script makes the ImageAverage structure based on previous
structures
%-------------------------------------------------------------------------
----
%

for i = 1:length(Data.AI);

   if isnan(Data.AI(i))
       i;

       continue
   end

   if Labels.YFPlabel == Data.ChannelLabels(i);
       YFP.ImageAverage = [YFP.ImageAverage Data.ImageAverage(i)];
   end

   if Labels.CFPlabel == Data.ChannelLabels(i);
       CFP.ImageAverage = [CFP.ImageAverage Data.ImageAverage(i)];
   end

   if Labels.RFPlabel == Data.ChannelLabels(i);
       RFP.ImageAverage = [RFP.ImageAverage Data.ImageAverage(i)];
   end

    if Labels.YFPlabel ~= Data.ChannelLabels(i) | Labels.CFPlabel ~=
Data.ChannelLabels(i) | Labels.RFPlabel ~= Data.ChannelLabels(i);
        %i; "~=" is the NOT EQUALS operator
    end

end
%-------------------------------------------------------------------------
----

%This section takes previously made structures and normalizes them
based on
%the image to image average and the RFP channel
CFP.AIMean = mean(CFP.AI);
YFP.AIMean = mean(YFP.AI);
RFP.AIMean = mean(RFP.AI);

YFP.ImageMultiplier = YFP.ImageAverage ./ RFP.ImageAverage;
CFP.ImageMultiplier = CFP.ImageAverage ./ RFP.ImageAverage;
```

```
YFP.ImageAdjust = YFP.ImageAverage ./ YFP.ImageMultiplier;
CFP.ImageAdjust = CFP.ImageAverage ./ CFP.ImageMultiplier;

% This is the normaliztion protocol for AImean
if CFP.AIMean/RFP.AIMean >= 0
    CFP.ImageNorm1 = CFP.AI ./ CFP.ImageMultiplier;
elseif CFP.AImean/RFP.AIMean <= 0
    CFP.ImageNorm1 = CFP.AI .* CFP.ImageMultiplier;
end

if YFP.AIMean/RFP.AIMean >= 0
    YFP.ImageNorm1 = YFP.AI ./ YFP.ImageMultiplier;
elseif YFP.AImean/RFP.AIMean <= 0
    YFP.ImageNorm1 = YFP.AI .* YFP.ImageMultiplier;
end

CFP.AINormmean = mean(CFP.ImageNorm1);
YFP.AINormmean = mean(YFP.ImageNorm1);

RFP.AIMean2 = RFP.ImageAverage ./ mean(RFP.AI);
CFP.AIMean2 = CFP.ImageAverage ./ mean(RFP.AI);
YFP.AIMean2 = YFP.ImageAverage ./ mean(RFP.AI);

%Creates the CYR.Norm2structures
if RFP.ImageAverage/RFP.AIMean >= 1
    RFP.AINorm2 = RFP.AI .* RFP.AIMean2;
elseif RFP.AIMean/RFP.AIMean <= 1
    RFP.AINorm2 = RFP.AI ./ RFP.AIMean2;
end

if CFP.ImageAverage/CFP.AINormmean >= 0
    CFP.AINorm2 = CFP.AI ./ CFP.AIMean2;
elseif CFP.AIMean/CFP.AINormmean <= 0
    CFP.AINorm2 = CFP.AI .* CFP.AIMean2;
end

if YFP.ImageAverage/YFP.AINormmean >= 0
    YFP.AINorm2 = YFP.AI ./ YFP.AIMean2;
elseif YFP.AIMean/YFP.AINormmean <= 0
    YFP.AINorm2 = YFP.AI .* YFP.AIMean2;
end

 windowsize=20;

%-------------------------------------------------------------------------
----
%This  section calculates RFP2, RFP3 and RFP4

    CFP.CFP2=zeros(size(CFP.AI,1)-windowsize);
    for ii=1:size(CFP.AI,2)-windowsize
```

```
        CFP.CFP2(ii)=mean(CFP.AI(ii:ii+windowsize));
    end
    YFP.YFP2=zeros(size(YFP.AI,1)-windowsize);
    for ii=1:size(YFP.AI,2)-windowsize
        YFP.YFP2(ii)=mean(YFP.AI(ii:ii+windowsize));
    end
    RFP.RFP2=zeros(size(RFP.AI,1)-windowsize);
    for ii=1:size(RFP.AI,2)-windowsize
        RFP.RFP2(ii)=mean(RFP.AI(ii:ii+windowsize));
    end

    CFP.CFP3=zeros(size(CFP.ImageNorm1,1)-windowsize);
    for ii=1:size(CFP.ImageNorm1,2)-windowsize
        CFP.CFP3(ii)=mean(CFP.ImageNorm1(ii:ii+windowsize));
    end
    YFP.YFP3=zeros(size(YFP.ImageNorm1,1)-windowsize);
    for ii=1:size(YFP.ImageNorm1,2)-windowsize
        YFP.YFP3(ii)=mean(YFP.ImageNorm1(ii:ii+windowsize));
    end
    RFP.RFP3=zeros(size(RFP.AI,1)-windowsize);
    for ii=1:size(RFP.AI,2)-windowsize
        RFP.RFP3(ii)=mean(RFP.AI(ii:ii+windowsize));
    end

    RFP.RFP4=zeros(size(RFP.AINorm2,1)-windowsize);
    for ii=1:size(RFP.AINorm2,2)-windowsize
        RFP.RFP4(ii)=mean(RFP.AINorm2(ii:ii+windowsize));
    end
    CFP.CFP4=zeros(size(CFP.AINorm2,1)-windowsize);
    for ii=1:size(CFP.AINorm2,2)-windowsize
        CFP.CFP4(ii)=mean(CFP.AINorm2(ii:ii+windowsize));
    end
    YFP.YFP4=zeros(size(YFP.AINorm2,1)-windowsize);
    for ii=1:size(YFP.AINorm2,2)-windowsize
        YFP.YFP4(ii)=mean(YFP.AINorm2(ii:ii+windowsize));
    end
%-------------------------------------------------------------------------
----

CFP.AIMeanArray(1,:) =
linspace(CFP.AINormmean,CFP.AINormmean,length(CFP.AINorm2));
YFP.AIMeanArray(1,:) =
linspace(YFP.AINormmean,YFP.AINormmean,length(YFP.AINorm2));
RFP.AIMeanArray(1,:) =
linspace(RFP.AIMean,RFP.AIMean,length(RFP.AINorm2));

% Calculating results and numerical cell/peroxisomes statistics

ResultsTable = [];
% Total number of peroxisomes in all images calculation
ResultsTable(1 , 2) = sum((RangeN(:,3))/3);
```

```
% Total number of cells in all images calculation
ResultsTable(2 , 2) = sum((RangeN(:,4))/3);
% Peroxisome/Cell ratio calculation
ResultsTable(3 , 2) = (ResultsTable(1 , 2)/ResultsTable(2 , 2));
% Number of images in this data set calculation
ResultsTable(4 , 2) = (length(RangeN(:,1))/3);

% This section is about performing the curve fitting analysis

[x]=[CFP.AI;CFP.ImageNorm1;CFP.AINorm2]';
[y]=[YFP.AI;YFP.ImageNorm1;YFP.AINorm2]';

for i = 1:3
  a(:,i) = polyfit(x(:,i),y(:,i),1);
end

% Calculate the sum of the squares of the residuals
% First calculate the values y_hat lying on the fitted straight line
and
% the corresponding residuals

for i = 1:3;
     y_hat(:,i) = polyval(a(:,i),x(:,i));
     residual(:,i) = y_hat(:,i)-y(:,i);
     residual_sq(:,i) = residual(:,i).*residual(:,i);
     Results.sum_sq_res(i) = sum(residual_sq(:,i));
     R_12 = corrcoef(x(:,i),y(:,i));
     Results.r_squared(i) = R_12(1,2)^2;
end

% Write the values for the slopes
Results.slope=a(1,:);
%disp(' ')
%fprintf('%s %d', 'The values for the slope are:', a(1,:))
% Write the values for the intercepts
Results.intercept=a(2,:);
%disp(' ')
%fprintf('%s %d', 'The values for the intercepts are:', a(2,:))
% calculate the mean of the x and y values
Results.mean_x = mean(x);
%disp(' ')
%fprintf('%s %d', 'The means of the x values are:', mean(x))
Results.mean_y = mean(y);
%disp(' ')
%fprintf('%s %d', 'The means of the y values are:', mean(y))

%Prints data from the RangeN array
disp(' ');
fprintf('%s %d', 'Total number of peroxisomes in all images:',
round(ResultsTable(1 , 2)))
disp(' ');
```

```
fprintf('%s %d', 'Total number of cells in all images:',
round(ResultsTable(2 , 2)))
disp(' ');
fprintf('%s %d', 'Peroxisome/Cell ratio:', ResultsTable(3 , 2))
disp(' ');
fprintf('%s %d', 'Number of images in this data set:', ResultsTable(4 ,
2))
disp(' ');

%First plot
%-------------------------------------------------------------------
----
%This part of the plot produces the scatterplot

figure
% %This gives an indication of the image to image variation in CFP
subplot(1,2,1)
    plot ((1:length(CFP.CFP2)),CFP.CFP2,'c',...
          (1:length(YFP.YFP2)),YFP.YFP2,'y',...

(1:length(RFP.RFP2)),RFP.RFP2,'r','LineWidth',2,'LineStyle','-')
%Syntax is (x1,y1,s1,x2,y2,s2,x3,y3,s3,...)
    title(['CYR Moving Average AI';'']);
    %title syntax: title(['First line';'Second line'])

    xlabel('Peroxisome Number')
    ylabel('Average Intensity')
    legend('CFP','YFP','RFP','Location','SouthOutside')
    axis square
%     mTextBox = uicontrol('style','text')
%     set(mTextBox,'String','Hello World')

   %set('LineWidth',3,'LineStyle',':')

% This plot shows the image to image average
subplot(1,2,2)
    plot (1:length(CFP.ImageAverage),CFP.ImageAverage,'c',...
          1:length(YFP.ImageAverage),YFP.ImageAverage,'y',...

1:length(RFP.ImageAverage),RFP.ImageAverage,'r','LineWidth',2,'LineStyl
e','-')
    title({'CYR Image to Image AI';''});
    xlabel('Peroxisome Number')
    ylabel('Image Mean Intensity')
    legend('CFP','YFP', 'RFP','Location','SouthOutside')
    axis square

figure
%This is a test plot of the image to image variation on all 3 channels
subplot(1,2,1)
    plot ((1:length(CFP.CFP2)),CFP.CFP3,'c',...
```

```
          (1:length(YFP.YFP2)),YFP.YFP3,'y',...

(1:length(RFP.RFP2)),RFP.RFP3,'r','LineWidth',2,'LineStyle','-')
%Syntax is (x1,y1,s1,x2,y2,s2,x3,y3,s3,...)
    title(['CYR 1st Normalized Moving AI';'']);
    xlabel('Peroxisome Number')
    ylabel('1st Normalized AI')
    legend('CFP','YFP','RFP','Location','SouthOutside')
    axis square

%This is a test plot of the image to image variation on all 3 channels
subplot(1,2,2)
    plot (1:length(CFP.ImageAdjust),CFP.ImageAdjust,'c',...
          1:length(YFP.ImageAdjust),YFP.ImageAdjust,'y',...

1:length(RFP.ImageAverage),RFP.ImageAverage,'r','LineWidth',2,'LineStyl
e','-')
    title(['CYR 1st Normalized Image AI';'']);
    xlabel('Peroxisome Number')
    ylabel('Image Mean 1st Normalized AI')
    legend('CFP','YFP', 'RFP','Location','SouthOutside')
    axis square

figure
% This plot shows the image to image corrected
subplot(1,2,1)
     plot((1:length(CFP.CFP4)),CFP.CFP4,'c-',...
          (1:length(YFP.YFP4)),YFP.YFP4,'y-',...
          (1:length(RFP.RFP4)),RFP.RFP4,'r-
','LineWidth',2,'LineStyle','-') %Syntax is
(x1,y1,s1,x2,y2,s2,x3,y3,s3,...)
    title(['CYR 2nd Normalized Moving AI';'']);
    xlabel('Peroxisome Number')
    ylabel('2nd Normalized AI')
    legend('CFP','YFP','RFP','Location','SouthOutside')
    axis square

subplot(1,2,2)
    plot(1:length(CFP.AIMeanArray),CFP.AIMeanArray,'c-',...
         1:length(YFP.AIMeanArray),YFP.AIMeanArray,'-y',...
         1:length(RFP.AIMeanArray),RFP.AIMeanArray,'r-
','LineWidth',2,'LineStyle','-');
    title(['CYR 2nd Normalized Image AI';'']);
    xlabel('Peroxisome Number')
    ylabel('Image Mean 2nd Normalized AI')
    legend('CFP','YFP','RFP','Location','SouthOutside')
    axis square

%-------------------------------------------------------------------------
----
%Second plot
```

```
%-------------------------------------------------------------------
----
%
myplot = figure(4)
%figure(4) = figure('PaperSize',[20.98 29.68]);

subplot(1,2,1)
    %plot(CFP.AI,YFP.AI,'o',CFP.AI,y_hat(:,1),'-')
    hold on
    dscatter(CFP.AI', YFP.AI')
    %dscatter(CFP.AI',(YFP.AI'),'plottype','contour')
    hold off
    %plot(x(:,1),y(:,1),'o',x(:,1),y_hat(:,1))
    title(['CFP vs YFP Scatterplot';'']);
    %xlim([0 100])
    %ylim([0 100])
    xlabel('CFP relative values')
    ylabel('YFP relative values')
    colorbar('Location','Southoutside','XTickLabel',...
             ['Lower Density',' ',' ','Higher Density'])
    %set(get(cbar_handle,'ylabel'),'string','My Title','fontsize',16)
    %legend('CFP/YFP','contours','Location','SouthOutside')
    %set(myplot,'Position',[100,100,500,500])
    axis square
    %[n,xout] = hist(CFP.AI, YFP.AI,40);

    %set(myPlot1,'Position',[100,100,500,500])


%This plots a histogram of the CFP/YFP ratio
subplot(1,2,2)
    hist (log(YFP.AI ./ CFP.AI), 50);
    title(['YFP/CFP histogram';'']);
    xlabel('Log of YFP/CFP')
    ylabel('Frequency')
    %set(myplot,'Position',[600,100,500,500])
    %legend('Data','Location','SouthOutside')
    %colorbar ('Location','Westoutside')
    axis square

figure(5)
subplot(1,2,1)


%plot(CFP.ImageNorm1,YFP.ImageNorm1,'o',CFP.ImageNorm1,y_hat(:,2),'-')
    hold on
    dscatter(CFP.ImageNorm1', YFP.ImageNorm1')
    %dscatter(CFP.ImageNorm1',(YFP.ImageNorm1'),'plottype','contour')
    hold off
    title(['CFP2 vs YFP2 1st Normalized Scatterplot';'']);
    %xlim([0 100])
```

```
    %ylim([0 100])
    xlabel('CFP2 relative values')
    ylabel('YFP2 relative values')
    colorbar('Location','Southoutside','XTickLabel',...
            ['Lower Density',' ',' ',' ','Higher Density'])
    %legend('Norm1','contours','Location','SouthOutside')
    axis square

%This plots a histogram of the CFP/YFP ratio
subplot(1,2,2)
    hist (log(YFP.ImageNorm1 ./ CFP.ImageNorm1), 50);
    title(['YFP2/CFP2 1s Normalized Histogram';'']);
    xlabel('Log of CFP2/YFP2')
    ylabel('Frequency')
    axis square

figure(6)
subplot(1,2,1)
    % plot(CFP.AINorm2,YFP.AINorm2,'o',CFP.AINorm2,y_hat(:,3),'-')
    hold on
    dscatter(CFP.AINorm2', YFP.AINorm2')
    %dscatter(CFP.AINorm2',(YFP.AINorm2'),'plottype','contour')
    hold off
    title(['CFP3 vs YFP3 2nd Normalized Scatterplot';'']);
    %xlim([0 100])
    %ylim([0 100])
    xlabel('CFP3 relative values')
    ylabel('YFP3 relative values')
    colorbar('Location','Southoutside','XTickLabel',...
            ['Lower Density',' ',' ','Higher Density'])
    %legend('Norm2','contours','Location','SouthOutside')
    axis square

%This plots a histogram of the CFP/YFP ratio
subplot(1,2,2)
    hist (log(YFP.AINorm2 ./ CFP.AINorm2), 50);
    title(['CFP3/YFP3 2nd Normalized Histogram';'']);
    xlabel('Log of CFP3/YFP3')
    ylabel('Frequency')
    axis square

%-------------------------------------------------------------------------
----
[n,AINorm2] = hist(log(YFP.AINorm2 ./ CFP.AINorm2), 50);
%This is the curve fitting part of
%-------------------------------------------------------------------------
----
%function [] = KJY15_Fit2(AINorm2,n)

%function KJY15_Fit2(AINorm2,n)
```

```
%Function syntax:
%function [output_parameter_list] = function_name(input_parameter_list)

%KJY15_FIT2    Create plot of datasets and fits
%   KJY15_FIT2(AINORM2,N)
%   Creates a plot, similar to the plot in the main distribution
fitting
%   window, using the data that you provide as input.  You can
%   apply this function to the same data you used with dfittool
%   or with different data.  You may want to edit the function to
%   customize the code and this help message.
%
%   Number of datasets:  1
%   Number of fits:  4

% This function was automatically generated on 27-Apr-2010 12:20:00
% AINorm2 = [];
% n = [];
%
% AINORM2 = AINorm2(:);
% n = n(:)

% Data from dataset "AINorm2 data (2 )":
%    Y = AINorm2
%    Frequency = n
%
 PlotData = [];
 %n = n/max(n);

% Force all inputs to be column vectors

% Set up figure to receive datasets and fits
f_ = [];
figure(7);
set(f_,'Units','Pixels','Position',[309 241 824 578.5]);
legh_ = []; legt_ = [];   % handles and text for legend
ax_ = newplot;
set(ax_,'Box','on');
hold on;

% --- Plot data originally in dataset "AINorm2 data (2 )"
t_ = ~isnan(AINorm2) & ~isnan(n);
Data_ = AINorm2(t_);
Freq_ = n(t_);
[F_,X_] = ecdf(Data_,'Function','cdf'...
               ,'freq',Freq_...
              );  % compute empirical cdf
Bin_.rule = 3;
Bin_.nbins = 50;
[C_,E_] = dfswitchyard('dfhistbins',Data_,[],Freq_,Bin_,F_,X_);
[N_,C_] = ecdfhist(F_,X_,'edges',E_); % empirical pdf from cdf
```

```
h_ = bar(C_,N_,'hist');
set(h_,'FaceColor','none','EdgeColor',[0.333333 0 0.666667],...
       'LineStyle','-', 'LineWidth',1);
xlabel('Data');
ylabel('Density')
legh_(end+1) = h_;
legt_[end+1] = 'YFP/CFP Histogram Data';

% Nudge axis limits beyond data limits
xlim_ = get(ax_,'XLim');
if all(isfinite(xlim_))
   xlim_ = xlim_ + [-1 1] * 0.01 * diff(xlim_);
   set(ax_,'XLim',xlim_)
end

x_ = linspace(xlim_(1),xlim_(2),100);

% --- Create fit "fit 1"

% Fit this distribution to get parameter values
t_ = ~isnan(AINorm2) & ~isnan(n);
Data_ = AINorm2(t_);
Freq_ = n(t_);
% To use parameter estimates from the original fit:
%     p_ = [ -0.04612222664033, 0.377542133208];
pargs_ = cell(1,2);
[pargs_[:]] = normfit(Data_, 0.05, [], Freq_);
p_ = [pargs_[:]];
y_ = normpdf(x_,p_(1), p_(2));
h_ = plot(x_,y_,'Color',[1 0 0],...
          'LineStyle','-', 'LineWidth',2,...
          'Marker','none', 'MarkerSize',6);
legh_(end+1) = h_;
legt_[end+1] = 'Normal Distribution';

PlotData.x1 = reshape(x_, 2, 50);
PlotData.y1 = reshape(y_, 2, 50);

% --- Create fit "fit 3"

%Fit this distribution to get parameter values
t_ = ~isnan(AINorm2) & ~isnan(n);
Data_ = AINorm2(t_);
Freq_ = n(t_);
%   To use parameter estimates from the original fit:
%     p_ = [ -0.05227557744762, 0.3571478053621, 19.1096594856];
p_ = mle(Data_, 'dist','tlocationscale', 'alpha',0.05, 'cens',[],
'freq',Freq_);  % Fit t location-scale distribution
y_ = pdf('tlocationscale',x_,p_(1), p_(2), p_(3));
h_ = plot(x_,y_,'Color',[0.666667 0.333333 0],...
          'LineStyle','-', 'LineWidth',2,...
```

```
          'Marker','none', 'MarkerSize',6);
legh_(end+1) = h_;
legt_[end+1] = 'T Location-Scale Fit';

PlotData.x3 = reshape(x_, 2, 50);
PlotData.y3 = reshape(y_, 2, 50);

% --- Create fit "fit 4"

t_ = ~isnan(Freq_) & ~isnan(n);
Data_ = AINorm2(t_);
Freq_ = n(t_);
y_ = ksdensity(Data_,x_,'kernel','normal',...
               'cens',[],'weight',Freq_,...
               'support','unbounded',...
               'width',0.25,...
               'function','pdf');
h_ = plot(x_,y_,'Color',[0.333333 0.333333 0.333333],...
          'LineStyle','-', 'LineWidth',2,...
          'Marker','none', 'MarkerSize',6);
legh_(end+1) = h_;
legt_[end+1] = 'Non-Parametric Fit';

hold off;
leginfo_ = ['Orientation', 'vertical', 'Location', 'SouthOutside'];
h_ = legend(ax_,legh_,legt_,leginfo_[:]);  % create legend
title(['YFP/CFP histogram data with fitted distributions';'']);
set(h_,'Interpreter','none');

PlotData.x4 = reshape(x_, 2, 50);
PlotData.y4 = reshape(y_, 2, 50);
%-------------------------------------------------------------------------
----
% This is the curve fitting graphs
%-------------------------------------------------------------------------
----
%This is a

%This is a temporary command as the KJY15_Fit run needs to be run for
this
%script to function
%run KJY15_Fit
run KJY15_Fit2

% PlotData.x = reshape(x_, 2, 50);
% PlotData.y = reshape(y_, 2, 50);

%This command removes the second row from the resized distribution
curves
PlotData.x1(2,:) = [];
PlotData.y1(2,:) = [];
```

```
%PlotData.x2(2,:) = [];
%PlotData.y2(2,:) = [];
PlotData.x3(2,:) = [];
PlotData.y3(2,:) = [];
PlotData.x4(2,:) = [];
PlotData.y4(2,:) = [];

%This command isn't quite finished. Need to have a better expression of
the
%scaling of the Freq_ array.

%PlotData.Freq_ = (Freq_ / max(Freq_));
 PlotData.Freq_ = ((Freq_ / max(Freq_)) * max(N_));
 PlotData.Freq_ = PlotData.Freq_';

PlotData.r1 = PlotData.y1 - PlotData.Freq_';
%PlotData.r2 = PlotData.y2 - PlotData.Freq_';
PlotData.r3 = PlotData.y3 - PlotData.Freq_';
PlotData.r4 = PlotData.y4 - PlotData.Freq_';

figure(8)
%figure(8) = figure('PaperSize',[20.98 29.68]);
%= figure('PaperSize',[20.98 29.68]);
subplot 131
plot (PlotData.x1(1,:),abs(PlotData.r1(1,:)),'-',...
      PlotData.x1(1,:),PlotData.Freq_,'-g',...
      PlotData.x1(1,:),PlotData.y1(1,:),'-
r','LineWidth',2,'LineStyle','-.')
  legend('Normal Distribution','Histogram
data','Difference','location','SouthOutside')
    title(['Normal Distribution';'']);
    xlabel('Data')
    ylabel('Scaled Intensity')
axis square

%Line style example txt: ('Marker, symbol, Color') For example,
plot(x,y,'-.or')

% subplot 122
% plot (PlotData.x2(1,:),PlotData.y2(1,:),'-',...
%       PlotData.x2(1,:),PlotData.Freq_,'-',... %Histogram data in life
format
%
PlotData.x2(1,:),abs(PlotData.r2(1,:)),'LineWidth',2,'LineStyle','-')
%   legend('Generalized Extreme Value','Histogram
data','Difference','location','SouthOutside')
%     title(['Generalized Extreme Value';'']);
%     xlabel('Data')
%     ylabel('Scaled Intensity')
%
%   axis square
```

```
%     (PlotData.x1(1,:),abs(PlotData.r1(1,:)),'-',...
%      PlotData.x1(1,:),PlotData.Freq_,'-g',...
%      PlotData.x1(1,:),PlotData.y1(1,:),'-
r','LineWidth',2,'LineStyle','-.')

%figure (9)
subplot 132
plot (PlotData.x3(1,:),abs(PlotData.y3(1,:)),'-r',...
      PlotData.x3(1,:),PlotData.Freq_,'-g',...
      PlotData.x3(1,:),abs(PlotData.r3(1,:)),'-
b','LineWidth',2,'LineStyle','-.')
  legend('T Location-Scale Fit','Histogram
data','Difference','location','SouthOutside')
    title(['T Location-Scale Fit';'']);
    xlabel('Data')
    ylabel('Scaled Intensity')
  axis square

subplot 133
plot (PlotData.x4(1,:),abs(PlotData.y4(1,:)),'-r',...
      PlotData.x4(1,:),PlotData.Freq_,'-g',...
      PlotData.x4(1,:),abs(PlotData.r4(1,:)),'-
b','LineWidth',2,'LineStyle','-.')
  legend('Non-Parametric Fit','Histogram
data','Difference','location','SouthOutside')
    title(['Non-Parametric Fit';'']);
    xlabel('Data')
    ylabel('Scaled Intensity')
  axis square

%This needs to be better as the best fitting distribution should
%numerically be the first one.
PlotData.R21 = sum(PlotData.r1.*PlotData.r1);
%PlotData.R22 = sum(PlotData.r2.*PlotData.r2);
PlotData.R23 = sum(PlotData.r3.*PlotData.r3);
PlotData.R24 = sum(PlotData.r4.*PlotData.r4);

PlotData.Results(1 , 1) = PlotData.R21;
%PlotData.Results(1 , 2) = PlotData.R22;
PlotData.Results(1 , 3) = PlotData.R23;
PlotData.Results(1 , 4) = PlotData.R24;

disp(' ');
fprintf('Sum of the squares of the residuals from the curve fitting =')
disp(' ');
fprintf('(The lower the number, the better the fit)')
PlotData.Results
%-------------------------------------------------------------------------
----
%This section summarises all the data into an Excel spreadsheet
```

```
%-------------------------------------------------------------------
----
%This script will make the cell reference that displays all the
information
%about a processed datasheet into an Excel file

ResultsTable2 = [];

%ResultsTable2[1,1] = [];
ResultsTable2[2,1] = 'Matlab summary script';
ResultsTable2[3,1] = 'File Name:';
ResultsTable2[3,2] = FileName;

%ResultsTable2[3,1] = [];
ResultsTable2[4,1] = 'The number of peroxisomes in all images =';
ResultsTable2[5,1] = 'The number of cells in all images =';
ResultsTable2[6,1] = 'Peroxisome/Cell ratio =';
ResultsTable2[7,1] = 'Number of images in this data set =';
ResultsTable2[4,2] = ResultsTable (1,2);
ResultsTable2[5,2] = ResultsTable (2,2);
ResultsTable2[6,2] = ResultsTable (3,2);
ResultsTable2[7,2] = ResultsTable (4,2);

ResultsTable2[8,1] = 'Sum of the squares of the residuals from the
curve fitting';
ResultsTable2[9,1] = 'Normal Distribution';
ResultsTable2[10,1] = 'Generalized Extreme Value';
ResultsTable2[11,1] = 'T Location-Scale Fit';
ResultsTable2[12,1] = 'Non-Parametric Fit';

ResultsTable2[9,2] = PlotData.Results(1,1);
ResultsTable2[10,2] = PlotData.Results(1,2);
ResultsTable2[11,2] = PlotData.Results(1,3);
ResultsTable2[12,2] = PlotData.Results(1,4);

%This part writes a summary Excel spreadsheet
a = FileName;
b = '_XLSheet';
FileName2 = sprintf('%s',a,b);
xlswrite(FileName2, ResultsTable2);
%-------------------------------------------------------------------
----

% %This is the curve fitting tool
 %dfittool
%
```